



Monitoring of Bioprocesses. Opportunities and challenges: Opportunities and Challenges

Pontius, Katrin

Publication date:
2019

Document Version
Publisher's PDF, also known as Version of record

[Link back to DTU Orbit](#)

Citation (APA):
Pontius, K. (2019). *Monitoring of Bioprocesses. Opportunities and challenges: Opportunities and Challenges*. Technical University of Denmark.

General rights

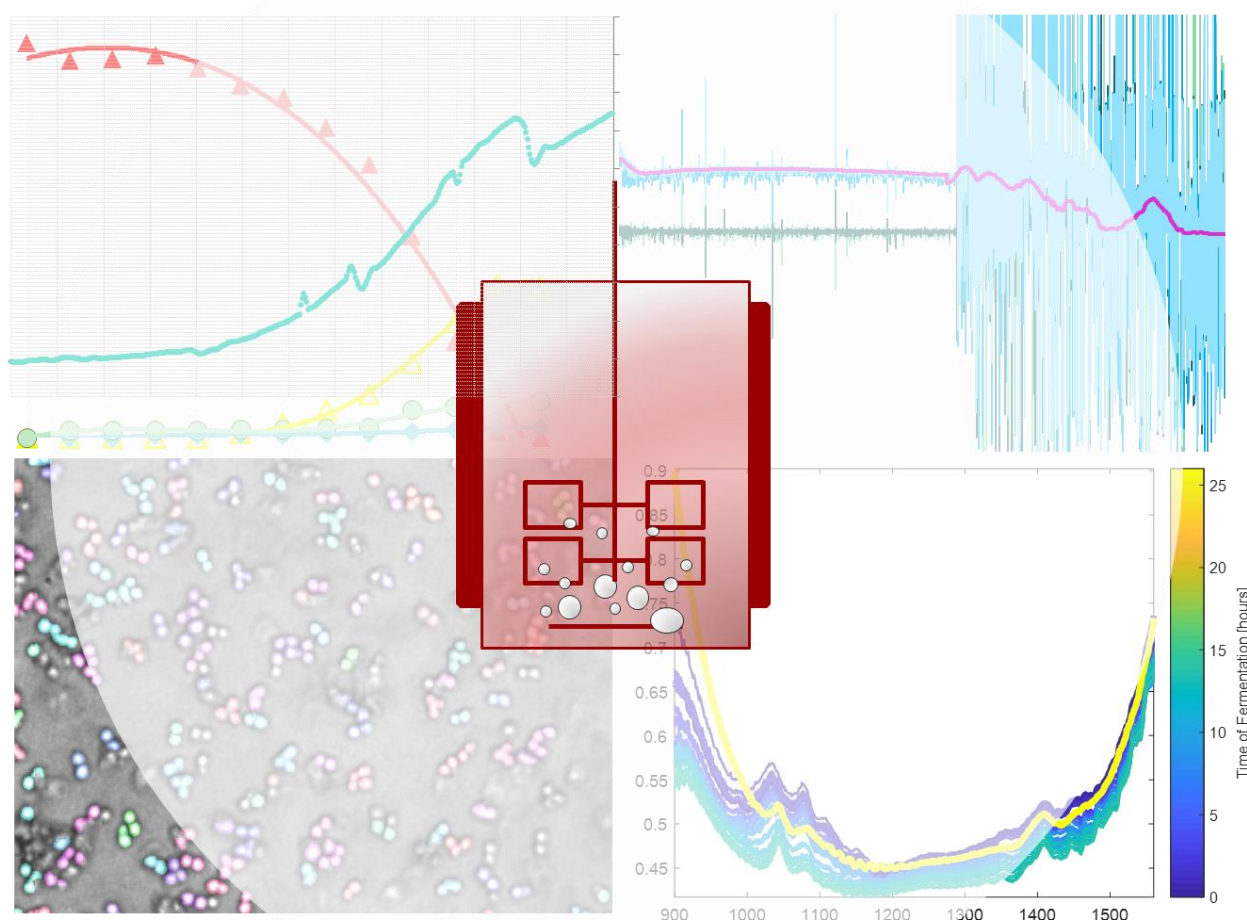
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Monitoring of Bioprocesses

Opportunities and Challenges



Katrin Pontius
PhD Thesis
May 2019

Monitoring of Bioprocesses Opportunities and Challenges

Ph.D. Thesis
Katrin Pontius

Process and System Engineering Centre
Department of Chemical and Biochemical Engineering
Technical University of Denmark

May 2019

Copyright ©: Katrin Pontius
May 2019

Address: Process and System Engineering centre (PROSYS)
Department of Chemical and Biochemical Engineering
Technical University of Denmark
Building 229
2800 Kgs. Lyngby
Denmark

Phone: +45 45252800

Web: www.kt.dtu.dk/forskning/prosys

Print: STEP

“In the middle of difficulty lies opportunity.”
Albert Einstein

Preface

The proposed thesis manuscript is submitted in order to obtain the PhD degree from the Technical University of Denmark (DTU). The presented work was accomplished in the period from December 2015 to May 2019 under the support and funding of BIOPRO SMV. The BIOPRO project has been carried out as a cooperation cluster between Danish universities and biotech industries aiming at making biotech production more sustainable and competitive throughout applied research. This PhD started under the supervision of Associate Professor Anna Eliasson Lantz and Ivan Hundebøl (DTU PILOT PLANT). However, it was physically located and accomplished at the Department of Chemical and Biochemical Engineering (DTU Kemiteknik), at the Process and Systems Engineering Centre (PROSYS) under Professor Krist V. Gernaey. Professor Gernaey has been involved as a co-supervisor during the first two years and took over the main supervision when Associate Professor Lantz decided to leave DTU in January 2018. Since then, the PhD was co-supervised by Assistant Professor Helena Junicke (DTU, PROSYS). Several project partners within the BIOPRO framework as well as external companies were involved during this PhD. The technologies under investigation were to a large extent developed in parallel to this work. An important project partner all through this PhD project was the University of Copenhagen, Department of Food Science (KU Food). The thesis as presented was realized through collaboration with:

- Marta Bevilacqua, KU Food, Copenhagen University, Denmark
- Jesper Bryde-Jacobsen, BIOPRO, Denmark
- Søren Friis, NLIR – Nonlinear Infrared Sensors, Denmark (SME started within BIOPRO)
- Trine Aabo Andersen, ParticleTech ApS, Denmark (SME started within f BIOPRO)
- Erik Spillum, BioSense Solution ApS, Denmark (SME started within BIOPRO)
- Paul Feng, NxPAS – The next generation photoacoustic systems, Denmark
- Anushka Vaibhav, Jobst Technologies GmbH, Germany
- Julius Netzer, aquila-biolabs GmbH, Germany

Acknowledgement

I would like to express my sincere gratitude to all my supervisors, Professor Krist V. Gernaey, Associate Professor Anna Eliasson-Lantz, Assistant Professor Helena Junicke and Ivan Hundebøl for the great opportunity to start and conduct my PhD at the Technical University of Denmark as part of the PROSYS research centre. It fulfilled my wish of experiencing a working environment 'abroad'. Thank you all very much for your support, trust, patience, understanding and critical discussions throughout the last 3.5 years. It has been a very special journey for me. I would like to give a special thank you to Krist V. Gernaey for fully adopting me into his group during my final year of PhD and for always being positive and easy-going. Another special thank you must be given to Helena Junicke, who just started her new position at DTU when taking over the co-supervision of my PhD. I count myself as very lucky having been part of this amazing group.

A great thank you to Marta Bevilacqua who has been like a supervisor for me all-through my PhD in the complex field of chemometrics.

This project was driven by cooperation and I would like to acknowledge all my project partners for a fruitful collaboration.

I would like address a special thank you to all my colleagues and friends from PROSYS. I totally appreciated the international, open, friendly and welcoming environment created by those fantastic people. I will miss you.

Another special thank you I would like to give to Gitte Læssøe who always helped me with a lot of friendliness and patience when needing to get through the administrative jungle.

To my family and friends, close and far, a thank you that comes from the deepest of my heart. Thank you for strengthening my soul, where ever I go.

Abstract

The monitoring of bioprocesses is a crucial issue and demand, not yet trivial outside the standard sensors comprising temperature, pH, dissolved oxygen and, in well-equipped laboratories and industries, off-gas analysis. The crucial process parameters such as biomass, substrate and product concentrations are rarely assessable on-line, a fact, that can compromise the reactor efficiency. Hence, this thesis addressed state-of-the art and novel technologies, commercially available but rather uncommonly considered as a monitoring strategy for fermentation processes. Within 4 individual chapters, the work investigated the monitoring of a yeast lab-scale fermentation process by means of:

- a biosensor designed as a small flow-through-cell for the monitoring of glucose;
- infrared (IR) spectroscopy combined with partial-least-squares (PLS) modeling for the monitoring of glucose, ethanol, glycerol, acetic acid, ammonium and phosphate;
- microscopic imaging and image analysis for the detection of growth supplemented with morphological characterization of the growing culture;
- a backscatter sensor for non-invasive monitoring of the microbial growth via the fermenter wall (glass);

It must be highlighted that all technologies could be applied on-line via a recirculation loop or non-invasively via the fermenter wall. The biosensor and the backscatter cell were found to be fast and easy to use as 'plug and play' devices, facilitating the monitoring of glucose and biomass, respectively, continuously on-line, during the entire course of the fermentation under study.

IR spectroscopy combined with PLS modelling yielded good results for the modelling of glucose and ethanol, while the monitoring of glycerol, acetic acid, phosphate and ammonium were compromised by *e.g.* high batch-to-batch variability, low concentration levels inside the broth, low IR-activity and indirect predictions. However, it is considered as a highly powerful approach on a long-term perspective. Imaging and image analysis was found to be an exciting new possibility, suggesting a totally novel, image based monitoring and control strategy. The technology in use was developed in parallel to this work and more research is needed to develop this approach into a final application.

The results achieved within this work were overall very promising. They allowed an integrated insight into the fermentation process under study while outlining the opportunities and challenges connected to the different technologies.

Dansk resume

Overvågningen af bioprocesser er et afgørende og efterspurgt foretagende hvortil der endnu ikke findes trivielle metoder, med undtagelse af standardsensorer til måling af temperatur, pH, opløst ilt samt gasanalyse af afkastgas i veludstyrede laboratorier. De afgørende procesparametre så som biomasse, substrat og produktkoncentrationer kan sjældent måles online, hvilket kan kompromittere nyttevirkningen af bioreaktoren. Derfor behandles der i denne afhandling en række state-of-the-art og nye teknologier der, på trods af kommerciel tilgængelighed, sjældent betragtes som overvågningsstrategi for fermenteringsprocesser. I 4 individuelle kapitler blev der i denne afhandling undersøgt hvorledes en fermenteringsproces i laboratorieskala med bage gær som mikroorganisme kunne overvåges ved hjælp af:

- En biosensor designet som en lille "flow-through"-celle til overvågning af glukose;
- Infrarød (IR) spektroskopi kombineret med partial-least-squares (PLS) modellering for overvågning af glukose, etanol, glycerol, eddikesyre, ammonium og fosfat;
- Mikroskopi og billedanalyse til påvisning af vækst suppleret med morfologisk karakterisering af vækstkulturen;
- en "backscatter" sensor til ikke-invasiv overvågning af den mikrobielle vækst via et vindue i fermentorvægen (glas);

Det skal understreges, at alle teknologier kunne anvendes online via en recirkulationssløjfe eller ikke-invasivt via fermentorvægen. Biosensoren og "backscatter" cellen viste sig at være hurtige samt brugervenlige som "plug and play" -enheder, der muliggjorde overvågningen af både glukose og biomasse kontinuerligt og online under hele fermentationsforløbet.

IR-spektroskopi kombineret med PLS-modellering gav gode resultater til modelleringen af glukose og etanol, medens overvågningen af glycerol, eddikesyre, fosfat og ammonium blev kompromitteret af f.eks. høj batch-til-batch-variabilitet, lave koncentrationer i fermenteringsvæsken, lav IR-aktivitet og indirekte prædiktioner. Det betragtes dog som en stærk tilgang på lang sigt. Billedanalyse viste sig at være en spændende ny mulighed, hvilket tyder på potentialet til at etablere en helt ny, billedbaseret overvågnings- og kontrolstrategi. Den anvendte teknologi blev udviklet parallelt med dette arbejde, og mere forskning er nødvendig for at udvikle denne tilgang til et færdigt produkt.

Resultaterne som blev opnået i dette arbejde var generelt meget lovende. Der blev opnået en indsigt i den undersøgte fermenteringsproces, samtidig med at muligheder og udfordringer forbundet med de forskellige teknologier blev belyst.

List of publications and conference contributions

Submitted Manuscript

Katrin Pontius, Giulia Praticò, Flemming H. Larsen, Thomas Skov, Nils Arneborg, Anna Eliasson Lantz, Marta Bevilacqua, *Fast measurement of phosphates and ammonium in fermentation-like media: feasibility study*, submitted to New Biotechnology, the 5th of February 2019.

Conference Contributions

Oral presentations

Katrin Pontius, Marta Bevilacqua, Krist V. Gernaey, Anna Eliasson Lantz, *Operational monitoring of phosphate and ammonium for an industrial fermentation process using infrared (IR) and near-infrared (NIR) spectroscopy*, 10th World Congress of Chemical Engineering held in Barcelona (Spain) from 1st to 5th October 2017

Katrin Pontius, Krist V. Gernaey Anna Eliasson Lantz, *Real-time monitoring of a fermentation process: linking yeast morphology to insulin production by image analysis*, Recent Advances in Fermentation technology (RAFT 12) held at the Hyatt Coconut Point, Bonita Springs (Ft. Myers), Florida, from 29th October to 1st November 2017

Poster presentations

Katrin Pontius, Marta Bevilacqua, Erik Spillum, Yuyang Feng, Helena Junicke, Anna Eliasson Lantz, Krist V. Gernaey, *Monitoring of a yeast insulin production process using Image Analysis and IR spectroscopy for integral process understanding and control*, 5th BioProScale Symposium, held in Berlin, Germany, from 20th to 22nd March 2018

Katrin Pontius, Marta Bevilacqua, Helena Junicke, Krist V. Gernaey, *Combining IR-spectroscopy and PLS modelling for advanced bio-process monitoring: simultaneous and rapid sensing of various components*, European Society of Biochemical Engineering Science (ESBES), held in Lisbon, Portugal from 9th to 12th September 2018

Contents

Preface	iii
Acknowledgement	iv
Abstract	v
Dansk resume	vi
List of publications and conference contributions	vii
Submitted Manuscript	vii
Conference Contributions	vii
List of abbreviations	xii
List of Symbols	xiii
Introduction to and scope of the thesis	1
Chapter 1	4
Monitoring of glucose applying a commercial biosensor system	4
Preface	4
1.1 Introduction	4
1.2 Materials and Methods	7
1.2.1 Glucose determination using a biosensor flow-through cell	7
1.2.2 Sensor Calibration.....	9
1.2.3 Off-line glucose measurements in yeast fermentation samples.....	10
1.2.4 On-line glucose monitoring during a yeast fermentation.....	11
1.2.5 Dissolved oxygen conversion inside fermentation samples.....	11
1.2.6 Storage conditions of the sensor	12
1.3 Results and Discussions	12
1.3.1 Investigations of sensor 1	12
1.3.1.1 Sensor calibration.....	12
1.3.1.2 Off-line glucose measurements on yeast fermentation samples	15
1.3.2 Continuous on-line glucose measurements during a 10 h yeast fermentation.....	17
1.3.3 Batch-to-batch variability and sensor stability	19
1.4 Conclusions	21
1.5 Supplementary Material	23
1.5.1 Figures	23
1.5.2 Tables.....	28

1.6 References.....	31
Chapter 2.....	33
Monitoring of several fermentation parameters applying infrared (IR) spectroscopy and partial -least -squares (PLS) modelling	33
Preface.....	33
2.1 Introduction.....	33
2.2 Materials and Methods.....	38
2.2.1 Data collection.....	38
2.2.1.1 Yeast fermentation processes	38
2.2.1.2 Design of synthetic samples	39
2.2.1.4 Preparation of synthetic samples.....	42
2.2.1.5 Spiked samples	42
2.2.3 Reference analysis.....	43
2.2.3.1 Glucose and metabolite quantification.....	43
2.2.2 Ammonium quantification	43
2.2.3 Phosphate quantification.....	43
2.2.3 Spectroscopic analysis.....	44
2.2.3.1 ATR-FT-IR (standard instrument)	44
2.2.3.2 NLIR- Nonlinear IR (prototype).....	44
2.2.4 Chemometric modelling	46
2.2.4.1 PLS calibration and PCA analysis of spectral data	47
2.3 Results and Discussions	49
2.3.1 Fermentation data.....	49
2.3.2 Spectral analysis.....	51
2.3.2.1 Pure component spectra	51
2.3.2.2 Fermentation spectra – raw spectra analysis	55
2.3.4 PLS modelling	57
2.3.4.1 Exploring the data and preprocessing	57
2.3.4.2 Several PLS-models – general observations	58
2.3.4.3 Detailed presentation of Model 3a	62
2.3.5 The effect of instrumentation (spectral quality) on the model performance.....	75
2.3.6 On-line MIR-spectroscopy.....	76
2.3.6.1 Influence of the flow-rate on on-line spectra collection.....	82

2.3.6.2 Pure component spectra acquired on-line.....	83
2.3.6.3 Workflow for PLS-model development under flow-conditions	85
2.4 Conclusions and Future Perspective	88
2.5 Supplementary Material.....	90
2.5.1 Tables.....	90
2.5.2 Figures.....	94
2.6 Refernces.....	98
 Chapter 3.....	 102
Monitoring of microbial growth and morphology dynamics by imaging and image analysis	102
Preface.....	102
3.1 Introduction.....	102
3.2 Materials and Methods.....	105
3.2.1 The oCelloScope detection principle.....	105
3.2.2 Detection of microbial growth.....	108
3.2.3 Morphological descriptors.....	109
3.2.4 Yeast fed-batch fermentation process.....	111
3.2.5 Investigating image acquisition and analysis – preliminary experiments.....	111
3.2.6 On-line image acquisition.....	112
3.2.7 Off-line image acquisition	113
3.2.8 Experimental set-up	113
3.3 Results and Discussion.....	116
3.3.1 The Fermentation Process.....	116
3.3.2 On-line fermentation images	117
3.3.3 Off-line fermentation images	120
3.3.4 Segmentation statistics.....	124
3.3.5 Microbial growth based on image analysis.....	128
3.3.6 Morphology dynamics based on image analysis.....	130
3.3.7 Performance of the bright spot feature.....	133
3.4 Conclusions and future perspectives	134
3.5 Supplementary Material.....	137
3.6 References.....	140

Chapter 4.....	142
Monitoring of microbial growth by the Cell-Growth-Quantifier	142
Preface.....	142
4.1 Introduction.....	142
4.2 Materials and Methods.....	143
4.2.1 Light scattering	143
4.2.2 The CGQ.....	144
4.2.3 Experimental conditions tested by the CGQ.....	146
4.2.3.1 CGQ set-up.....	146
4.2.3.2 Operational conditions under investigation	146
4.3.2.2 Fermentation conditions under investigation	147
4.3 Results and Discussions	148
4.3.1 Influence of stirrer speed and aeration rate on the sensor baseline signal.....	148
4.3.1.1 Water	148
4.3.1.2 YP medium.....	150
4.3.1.3 YP medium containing particles.....	151
4.3.1.4 Influence of stirrer speed and aeration rate – summary	153
4.3.2 Monitoring of Biomass by the CGQ.....	154
4.3.2.1 Fermentation 1	154
4.3.2.2 Fermentation 2	156
4.4 Conclusions	158
4.5 Supplementary Material.....	159
4.6 References.....	160
Final Conclusions.....	161
Appendix 1.....	164

List of abbreviations

ATR	attenuated total reflectance
CENPK	laboratory yeast strain
CGQ	Cell Growth Quantifier (aquila biolabs, Germany)
GOx	Glucose Oxidase
dd water	distilled deionized water
DO	dissolved oxygen
DOE	design of experiments
DW	dry weight
FT-IR	Fourier-Transform Infrared Spectroscopy
IR	Infrared
LV	latent variable
Milli Q	dd water (purified water by means of Millipore equipment)
MIR	Mid Infrared
MIRS	Mid Infrared Spectroscopy
NIR	Near Infrared
NIRS	Near Infrared Spectroscopy
NLIR	Nonlinear Infrared Spectroscopy
OD	optical density
PAT	process analytical technology
PCA	principle component analysis
PLS	partial -least -squares
r	correlation coefficient
R ²	squared correlation coefficient
RMSE(Cal)	root mean square error of calibration
RMSE(CV)	root mean square error of cross validation
RMSEP	root mean square error of prediction
SIC	insulin precursor producing yeast strain
SNR	signal to noise ratio
SNV	standard normal variate
st. dev.	standard deviation
VIP	very important variable
YP	medium containing yeast extract (Y) and peptone (P)
YPD	medium containing yeast extract (Y), peptone (P) and dextrose (D)

List of Symbols

c	speed of light
E	Energy
f	Frequency
h	Planck's constant
I	light intensity
I_0	incoming light intensity
N	total number of samples
r	correlation coefficient
R^2	squared correlation coefficient
\mathbf{X}	matrix (spectral data)
\mathbf{y}	vector (of reference concentration values)
y	a reference value
\hat{y}	a predicted value
α	absorption coefficient
λ	wavelength
τ	turbidity coefficient
ν	Wavenumber

Introduction to and scope of the thesis

Bioprocess manufacturing has played a key role in food, pharma and the chemical industry for the last 50 years. The producing core of any biotech industry is the fermentation process itself. It is often considered as the most complex unit operation within bio-manufacturing. Evidently, established industrial processes meet the productivity and quality demands raised from authorities and the market. However, proven process strategies might be outcompeted soon by increasing market and quality demands not sufficiently addressed in the existing industrial bioprocess framework. Modern industrial processing in the 21st century poses paradigms such as circular economy and sustainability to any industry that can only be met by improved and optimized monitoring and control strategies, aiming at increasing productivity while minimizing waste.

In contrast to this ideal, yet crucial for future bioprocessing, fermentation process design is still generally challenged by a lack of tools for process monitoring and control. Maximum reactor efficiency through profound process understanding and targeted control strategies can only be established on the basis of targeted monitoring strategies. However, control actions – frequently based on experience instead of process data – are still executed manually to a high degree. Entire batches might be discarded, resulting in a significant financial loss and environmental burden due to a monitoring deficit of crucial process parameters. Outside the well-established and trusted sensor systems for pH, temperature, dissolved oxygen, and in well-equipped laboratories and industries, off-gas analysis, real-time measurements of biomass, substrate and product concentrations are rare and possible solutions are highly challenged by medium and process complexity. The process analytical technology (PAT) framework published in 2004 [1] has addressed this crucial demand of process understanding through appropriate monitoring strategies and both, research and industry, are focusing increasingly on providing generic solutions. Besides, a new driving force has been created by the vision of Industry 4.0 [2] announced as one of the key initiatives in 2011 by the German federal government. The convergence of industrial production, process information and communication can only be achieved on the basis of valid monitoring strategies.

Recent advances in several technologies, in- and outside the field of biotechnology, have led to integrated solutions applicable to bioreactors. However, they are broadly overlooked. Bioprocess analysis is still predominately performed by means of the established standard methods such as optical density measurements and chromatographic analysis, facilitating the detection of the microbial growth and substrate and product concentration levels off-line. However, they cannot be considered as on-line monitoring strategies as their procedure requires time and results are usually only available after the process has finished making precise real-time control actions impossible. Yet, no generic solution meeting the vision of a bioprocess sensor, often pronounced as a probe comparable to the standard pH and dissolved oxygen sensor, is available. Such an ideal sensor is often conceptualized by the following: It can be inserted into the bioreactor via a standard port; it can be autoclaved; it is not affected by the process operations such as stirring and aeration and it

does not interfere with the process; it covers the detection of the entire concentration range relevant in bioprocesses and delivers accurate and continuous data on-screen. Ideally, the crucial process parameters, outlining any fermentation process, such as biomass, substrate and product concentration levels, can be measured by means of one and the same generic probe. However, as mentioned before, such a probe is not yet available. Fortunately, commercially available advanced technologies are facilitating the monitoring of individual fermentation parameters in integrated designs, and they have the potential to set new standards in the field of fermentation monitoring.

This thesis presents state-of-the-art and novel technologies, commercially available, but rather uncommonly used for the monitoring of fermentation processes. The technologies are all exemplarily applied to a *Saccharomyces cerevisiae* (fed-batch), lab scale fermentation.

Yeast is one of the most established workhorses in biotechnology, in both research and industrial application. In the last decades, it has been developed into a huge family of application optimized strains. The strain CEN.PK113-7D was the model microorganism within this work. The CEN.PK strains were developed for studying metabolic fluxes and gene expression and are widely used in fundamental and applied research. The genome of the strain CEN.PK113-7D was recently sequenced [3], [4].

The thesis is divided into 4 experimental chapters, presenting one technology each. Each technology can be applied on-line to the bioreactor, by means of a recirculation loop. The chapters can be read individually and are structured in the form of a manuscript, comprising an introduction, followed by the materials and methods, concluding with the results and discussions.

Chapter 1 starts with the monitoring of the essential parameter glucose, the substrate for various fermentation processes. Monitoring of glucose was conducted by means of an electrochemical biosensor (B.LV5, Jobst Technologies GmbH, Germany) designed as a flow-through-cell. Originally developed for pre-clinical studies in the diabetes field, the biosensor's design allows the connection to a bioreactor, as well. Hence, when I met the company Jobst technology during a conference stay in March 2018, I decided to integrate their system as a case study into this thesis.

Chapter 2 extends the monitoring of glucose to the monitoring of the yeast specific process performance parameters, ethanol, glycerol and acetic acid. The parameters ammonium and phosphate were included additionally to continue earlier work, presented in appendix 1. The monitoring of the 6 analytes glucose, ethanol, glycerol, acetic acid, ammonium and phosphate was conducted by means of a prototype of a novel infrared (IR) technology, the so-called nonlinear IR (NLIR, Nonlinear Infrared Sensors, Denmark), combined with partial -least -squares (PLS) modelling. It was performed in close collaboration with Copenhagen University, KU Food, and must be considered as the most labor and time intense chapter within this thesis. It outlines all issues connected to the highly interdisciplinary and complex field of IR spectroscopy and chemometric modelling.

Finally, the monitoring of the biomass itself is addressed in chapter 3 and 4.

In chapter 3, monitoring of biomass was enabled by recent advances in microscopy-based image analysis, exemplarily demonstrated by the oCelloScope instrument (BioSense Solutions Aps, Denmark). The technology facilitates the detection of microbial growth supplemented with

morphology analysis over the fermentation time, by imaging and image analysis. Literally observing the fermentation process via the process window created by the oCelloScope instrument, can be considered as a totally novel and unexplored approach.

Chapter 4 presents the non-invasive monitoring of microbial growth by means of the so-called Cell Growth Quantifier (CGQBIOR, aquila biolabs, Germany). The sensor applies backscatter technology via the fermenter wall (glass), or a small glass window. The case study was implemented into this work when I met the company aquila biolabs during a conference stay in March 2018. The technology was originally developed for the application on shake flasks, and no detailed data obtained by the application of the device to fermenter vessels was available by that time.

The different chapters introduce the respective technologies and investigate their performance with respect to their application as a monitoring tool for the bioprocess area.

References

- [1] FDA, "Guidance for Industry - PAT A Framework for Innovative Pharmaceutical Development, Manufacturing, and Quality Assurance," no. September, p. 1, 2004.
- [2] M. Hermann, T. Pentek, and B. Otto, "Design Principles for Industrie 4.0 Scenarios," in *2016 49th Hawaii International Conference on System Sciences (HICSS)*, 2016, pp. 3928–3937.
- [3] J. Otero *et al.*, "Whole genome sequencing of *Saccharomyces cerevisiae*: from genotype to phenotype for improved metabolic engineering applications," *BMC Genomics*, vol. 11, no. 1, p. 723, 2010.
- [4] J. F. Nijkamp *et al.*, "De novo sequencing, assembly and analysis of the genome of the laboratory strain *Saccharomyces cerevisiae* CEN.PK113-7D, a model for modern industrial biotechnology," *Microb. Cell Fact.*, vol. 11, no. 1, p. 36, 2012.

Chapter 1

Monitoring of glucose applying a commercial biosensor system

Preface

The following chapter presents the possibility of on-line and fast on-line monitoring of the essential parameter glucose by means of a commercially available glucose biosensor. The quantification of glucose is probably most advanced in the field of healthcare applications for the diabetes sector. For the detection of blood sugar, biosensors play an important role and have been thoroughly studied. However, glucose monitoring is of high relevance in various fermentation processes, too, and no general tools are yet implemented for this purpose. Nowadays, sophisticated biosensor solutions are commercially available and integrated in designs suitable for the application to fermentation processes. Yet, outside China [1], they have been broadly neglected for the monitoring and control of fermentation processes. When I met the German company Jobst Technologies GmbH during my visit of the 5th BioProScale Symposium, hold in Berlin, Germany, in March 2018, I decided to test their sensor with respect to fermentation monitoring. It must be highlighted that the sensor under investigation was able to detect glucose concentrations up to 150 mM in cell- containing samples, while most glucose biosensors described in literature facilitate glucose determination only up to 25 mM, mostly in cell-free samples. Hence, the sensor presents an interesting small and affordable tool for continuous glucose monitoring during the entire fermentation process. The work is considered for submission to the journal New Biotechnology.

The inspiration for this research application was given to me by my colleague Daria Semenova, who conducted her PhD on the topic of biosensor design and optimization. I would like to acknowledge Daria for her inspiration, support and critical feedback on the experiments conducted with the sensor. I would also like to acknowledge the people from Jobst Technologies (Freiburg, Germany) for their feedback on this chapter and manuscript to be submitted as well as for providing helpful information regarding their sensor platform. Finally, I'm grateful to Frans v. d. Berg, University of Copenhagen, Department of Food Science, for providing the Matlab script for smoothening the signal of the second sensor.

1.1 Introduction

The concept of process analytical technology (PAT) published by the Food and Drug Administration (FDA) in 2004 [2] must be seen as a revolutionary milestone not only in the pharmaceutical industry. The idea emphasized within the PAT framework, building quality through deep process understanding enabling data driven control of the critical quality attributes, strongly applies to the biotech industry as well. Although fermentations are often the most complex step within bio-manufacturing, until today, fermentation production reactors are generally rather sparsely instrumented and typically only involve standard sensors such as pH, temperature and dissolved

oxygen [3], [4]. Real-time information of the critical process parameters is often lacking and thus, control actions - frequently implemented manually, not automated - are based on experience instead of process data. Data collection is time and resource expensive and thus, most data is obtained in research and development (R&D), as higher costs are more acceptable there compared with industrial processing. Fact is, in R&D as well as in industrial production, analysis determining cell performance and physiology reflected in substrate, metabolite and product levels in the fermentation medium is mostly performed off-line after the process has finished. This causes a delay of at least several hours. Assaying the process relevant components is generally done by means of specific assay kits and chromatographic analysis. These methods are labor and time intensive and as such not suitable for prompt analysis or on-line monitoring strategies. Some fully automated systems for multicomponent analysis, even applying biosensor technology as emphasized in this study, have been developed for rapid off-line analysis reducing time and operational errors. Some of these analyzers even offer an on-line solution. However, they are voluminous and highly expensive and therefore not considered by every laboratory (e.g. Cedex Bio from Roche Diagnostics GmbH, the Biochemical Analyzer series from Yellow Springs Instruments (YSI, USA), the BioProfile series from nova biomedical, the analyzer series from SBA (China), as well as BioPAT® Trace and BioPAT® Multi Trace sold by Satorius).

Glucose is a major carbon and energy source in the fermentation industry and as such, evidently, the monitoring and control of glucose concentration levels during fermentation processes is beneficial to any feeding strategy, optimizing biomass production itself as well as the production of metabolites such as amino acids, alcohols, peptides and proteins. As mentioned before, also the important analyte glucose is mostly quantified in the fermentation broth by enzymatic assay kits or HPLC analysis taking at least 20 min of time. Note that, samples collected during the process are usually analyzed after the process for practical reasons as equipment scheduling and operational shifts. For 'quick and dirty' determination of glucose levels within a minute, single use test stripes giving a rough idea of the glucose concentration present based on a color development scale, can be used (e.g. Medi-Test Glucose from Machery-Nagel, originally applied for glucose determination in urine). However, those are neither suitable for continuous glucose monitoring and control strategies nor precise results. Real-time glucose monitoring systems are very rare and predominantly dedicated to the clinical use in the diabetes sector[5]. Although virtually, all glucose biosensors have been developed for diabetes health care applications, they also, or precisely therefor, provide all features desired for bioprocess monitoring: Per definition, biosensors are operated without the need of reagent addition and as such, they are simple to handle, fast, accurate and reliable. Besides, they are small and can be produced cheap in bulk quantities based on screen printing and thin film technology [4], [6]. The principle of enzyme based (mostly Glucose Oxidase, GOx) glucose biosensors was introduced more than 50 years ago by Clark and Lyons in 1962 [7] and since then, the biosensor field has grown enormously. An introduction to the field of electrochemical enzymatic and non-enzymatic biosensors, to their principle, challenges, fabrication and advances can be found in [8], [9], and [10]. Recent advances refer e.g. to non-enzymatic biosensors, better known as 4th generation biosensors. Non-enzymatic biosensors are based on advances in material science leading towards sophisticated electrode materials mimicking enzyme activity. Developed and designed for the determination of glucose in blood samples certainly showing comparable matrix complexity as fermentation broth, they appear to be highly suitable for analysis in the field of fermentation science. Interestingly, application of first- and second-generation biosensors was already proposed and

described in the late 80ies / early 90ies. Already in 1987 an autoclavable glucose biosensor for online monitoring of glucose in fermentation media was proposed [11] and the concept was refined into a prototype in 1995 [12]. With the prototype, glucose was controlled manually at a level of 0.44 g/l (2.44 mM) during an *E. coli* fed-batch fermentation for 2 h based on the sensor output demonstrating its potential for online bioprocess monitoring and control. Besides, continuous glucose monitoring was successfully achieved by the prototype during low (max. 2.2 g cell dry weight/liter) and high cell density (max. 85 g cell dry weight /liter) fed-batch fermentation for 16 and 27 h, respectively. The calibration in the described application ranged up to 23 mM glucose in cell-free samples. Grunding and Ch. Krabisch [13] published the design and use of a 2nd generation glucose biosensor for the determination of glucose in fermentation samples in a range of 0.1 – 20 mM, stating long term stability and usability of more than 12 weeks (frequently determined on fermentation samples of 10 g/l cell dry weight). J. Rishpon *et al* (1990) demonstrated in-situ glucose monitoring in yeast fermentation broth based on a 'sandwich GOx electrode' enabling glucose quantification in a range of 50 – 200 mM throughout a 3 h fermentation process in a 30 ml reaction vessel [14]. White *et al.* (1996) [15] suggested a screen printed glucose biosensor integrated in a flow injection system showing long-term stability over a 7 day period operating the system continuously with buffer, injecting a 10 mM glucose solution every 30 min. Hence, the interest and believe in the biosensor concept suitable for continuous glucose monitoring in fermentation processes was demonstrated around 25 years ago.

Contrarily, although the number of glucose biosensor relevant articles published in the last 15 years has maintained an increasing trend [8], their application published on bioprocesses remains rare and shows a gap between the 90is and the recent years practice. Publications reviewing the application of biosensors since 2014 proof the wide disregard of studying biosensors as a fermentation monitoring tool [1], [16], [17],[18]. The stagnancy, especially in fermentation application might be due to challenges in long term stability of the enzyme, the narrow detection range (usually up to only 25 mM due to relevant blood sugar levels in diabetes patients), sterility concerns or simply because no satisfying, ready to use, commercial solutions have been available applicable to fermentation processes (as health care surly demands a different integration than the biotech environment). The recent innovations in glucose biosensor technology and their application on fermentation processes include miniaturized biosensor arrays for simultaneous detection of up to 4 components [5], [19], their fabrication as a flow through unit [20] as well as enzyme-free glucose sensors mimicking enzyme specificity [21]. They all have in common a glucose detection range from 0.05 - 25 mM and are mostly described as a subject of scientific interest rather than an application as an on-line monitoring tool.

Within this study, I want to demonstrate the reliable use of a commercially available biosensor solution meeting the needs for glucose monitoring in fermentation processes. The sensor applies the principle of first generation glucose biosensors facilitating the detection of glucose within an extended glucose range of 1 – 150 mM, applying a segmented calibration curve for low (1 -50 mM) and high (50 – 150 mM) glucose concentration levels. It combines recent advances such as multi array design and flow through integration. It is of small size and has a reasonable price. The sensor performance was studied in different media, starting from the recommended acetate buffer system over complex YPD fermentation medium to finally real, cell-containing samples of a yeast fermentation. Finally, the sensor was applied continuously for 10 h directly connected to a 2 L lab-

scale fermenter via a recirculation loop during a yeast fed-batch cultivation. The sensor facilitated fast and easy data collection, both off-line and on-line. Thus, it supports time and resource efficient optimization and control of fermentation processes realizing maximum reactor efficiency.

1.2 Materials and Methods

Two biosensors of the same type were tested. The first one was in particular studied regarding calibration in different media, off-line analysis of fermentation samples, both cell-containing and cell-free, functionality over time (calibration before and after the fermentation samples) and finally performance of the sensor after a 3 month storage period. In order to investigate batch-to-batch variations of the biosensor, the second sensor was investigated regarding calibration before and after continuous use over a 10 h period during a fed-batch fermentation, testing the sensor in a true on-line set up, as well as its calibration in different media. The experiments conducted with sensor 1 and 2 are summarized in [Table 1.1](#).

Table 1.1: Experimental overview of sensor 1 and 2

Sensor	Experimental investigation
Sensor 1	Investigation of the sensor system 1) Calibration in different media, applying flow and non-flow conditions; 2) Off-line analysis of fermentation samples (cell-free and cell-containing); 3) Recalibration after usage in 1) and 2); 4) Storage stability after calibration, measurement of off-line fermentation samples and a storage period of 3 month;
Sensor 2	Batch to batch variability of the sensors 1) Calibration in different media, applying flow; 2) Continuous on-line measurements over the course of a 10 h fermentation; 3) Recalibration after usage in 1) and 2);

1.2.1 Glucose determination using a biosensor flow-through cell

The setup for glucose monitoring consisted of the biosensor B.LV5 (extended range sensor, [Figure 1.1A](#)), the potentiostat (SIX transmitter, [Figure 1.1 B](#)) with a custom connection for fitting the biosensor, as well as the software bioMON used for operating the SIX transmitter. All components are provided by Jobst Technologies GmbH (Freiburg, Germany). The components are comparable with 'plug and play' devices, in which the biosensor is connected to the SIX transmitter which is connected to the computer via a USB connection ([Figure 1.1 B, C](#)). The biosensor itself is designed as a 1 µl flow-through-cell continued in outer tubing (0.5 mm ID) representing inlet and outlet of the electrochemical cell, ending in luer fittings for fast and easy connection of additional tubing.

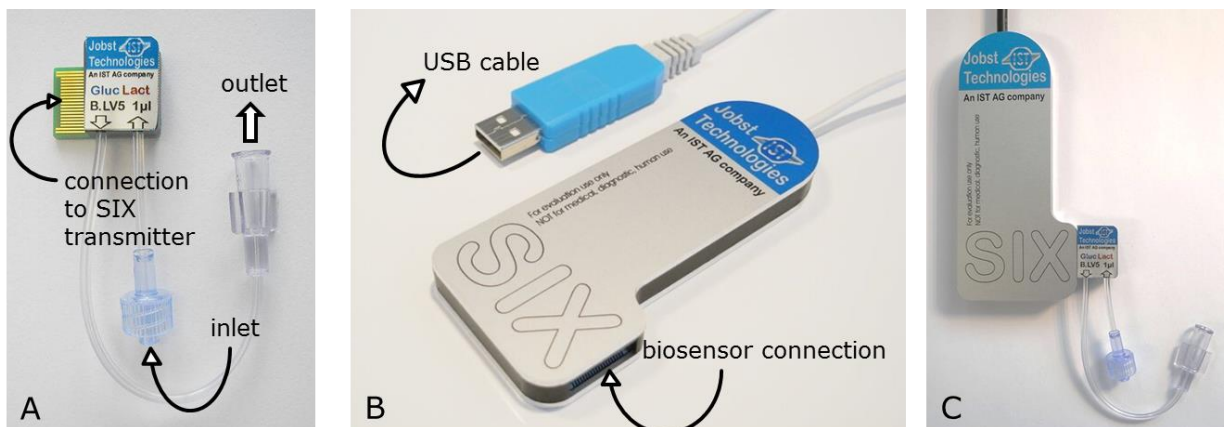


Figure 1.1: A) The biosensor flow-through-cell with connection to the SIX transmitter and luer fittings as inlet and outlet. B) The SIX transmitter with a free biosensor connection and USB cable. C) The biosensor connected to the SIX transmitter.

In order to have the sample flowing through the biosensor, a pump (Ismatec Reglo ICC, Cole-Parmer, UK) was connected via a piece of tubing (Ismatec, TYGON S3TM, E-LFL, ID 1,52 mm) and the respective luer connector (see Figure 1.1 A) to the inlet of the sensor.

The biosensor applies the principle of first generation glucose biosensors: A thin layer of the enzyme glucose oxidase (GOx) comprising the enzyme-bound cofactor FAD/FADH₂ is entrapped into a hydro gel membrane and placed on top of an H₂O₂ sensitive electrode. Glucose is enzymatically oxidized to gluconic acid, generating H₂O₂. The amount of H₂O₂ produced is proportional to the glucose concentration present and detected by anodic oxidation at a noble metal (Pt) electrode. The oxidation of glucose at the electrode produces an amperometric signal (electric current) which (in steady state) can be correlated with the respective glucose concentration. The redox reactions are summarized in eq. (1.1-1.3). More information about the sensor under study, its properties and manufacturing can be found in reference [6] and [5].



The biosensor, or the electrochemical cell respectively, consists of a Pt-working electrode, a Pt-counter electrode and an internal Ag/AgCl pseudoreference electrode. Oxidation of H₂O₂ is performed on the Pt-working anode at + 450 mV versus internal Ag/AgCl. Duplicates of each measurement were achieved by the multi array design of the sensor consisting of two working electrodes for glucose detection. Consequently, the system provided an internal duplicate for each sample measured. Independent of the presence of glucose in the medium, the medium itself results in a very small background current. In order to eliminate signal variations due to the background current or resulting from possible disturbances like air bubbles, for instance, the electrochemical cell provides two blank ('clean', non-enzyme coated) Pt-working electrodes, one for each glucose electrode. Subtracting the blank current from the glucose signal current led to the final measurement value. Conventionally, the calibration of an electrochemical biosensor is done giving the resulting current at a defined time point (often at 90 % conversion, not at steady state) as a function of glucose concentration. As in this study, the stabilization time for each sample was found to be quite different depending on the glucose concentration present, the medium used and the cell concentration, the

final current value was obtained by averaging the last 10 data points of the current in steady state. This holds for every current value obtained from glucose measurements with the biosensor within this work.

Glucose samples were quantified under non- flow and flow conditions, in the latter case a constant flow rate of 0.2 ml/min was applied. Under non-flow conditions, the sample was pumped into the biosensor, the pump was stopped and the measurement was started with the sample standing still in the measurement cell. When the measurement was finished, indicated by a final constant current, the sample was pumped out and the next sample was pumped into the sensor flushing the flow-through cell with approximately 100 μ l of new sample before stopping the pump and starting the measurement of the new sample. The procedure was equal for samples measured under flow, with the difference that the pump was kept running when starting the glucose measurement. Samples used for calibration of the sensor were pumped in and out of the sensor one after each other, a small volume of air in between samples (indication the end of the old / the beginning of the new sample). In case of measuring the glucose concentration in the fermentation samples, the sensor was stored in buffer in between the samples of different time points.

1.2.2 Sensor Calibration

According to the biosensor info sheet, the linear detection range of the sensor is within 0.05 mM and 60 mM glucose (applicable in a pH range of 5-9) and the sensor comes pre-calibrated from the company (calibration at company site performed in acetate buffer at 32 °C). However, in order to study the behavior of the biosensor under different conditions aiming at monitoring a yeast fermentation process, it was calibrated using both, the recommended acetate buffer solution as well as the fermentation medium used for yeast cultivation. The glucose calibration range was chosen to be within 1 to 150 mM, challenging the sensor system with respect to the upper detection limit particularly. Besides, aiming at on-line quantification without the need of sample dilution or preparation, the upper limit refers to the initial glucose concentration present in the reference fermentation process. The lower limit of calibration was chosen to be 1 mM as considered to be the lowest glucose concentration accurately measured by the reference method (HPLC). As the Ag/AgCl reference electrode depended on a sufficient concentration of chloride ions in the solution (recommended is around 110 mM in the biosensor manual) and the chloride concentration estimated in the medium was only about 5 mM, a third calibration solution was studied containing YPD medium and additional 9 g/l (154 mM) NaCl. This NaCl concentration was chosen, as saline solution (9 g/l NaCl in water) is commonly used as an isotonic non-nutritional dilution solution when handling microbial cells and thus considered to be harmless to the fermentation samples.

Finally, the calibration curve obtained in YPD medium was subsequently used to determine the glucose concentration in the samples of the yeast fermentation process off-line and on-line. The calibration was performed at room temperature, but the calibration samples were preheated to 30 °C according to the process temperature of the fermentation. In all calibration solutions the pH was adjusted to the operational pH during the fermentation (pH 6) by using the same acid (2 M H₂SO₄) as used for pH control during the fermentation process. The glucose concentrations of all solutions were determined in duplicates by HPLC measurements as reference method.

When referring to 'buffer' within this work, it was consistently the acetate buffer as follows: Per liter of purified water, it contained 0.313 g potassium chloride, 5.443 g sodium acetate trihydrate, 5.669 g sodium chloride, 0.014 g sodium phosphate monobasic monohydrate, 0.114 g sodium phosphate dibasic dihydrate, 0.122 g magnesium chloride dehydrate and 1ml Proclin 300 to avoid microbial activity. All ingredients were purchased from Sigma, USA.

YPD medium is a classical complex medium used for yeast cultivations. It contained per liter of water 10 g yeast extract (Y, Merck France, NaCl < 5 %), 20 g peptone (P, Merck Mexico, NaCl < 3 %) and 20 g glucose (dextrose, D, Macron, USA). In case of NaCl addition, 9 g/l NaCl (purchased from Sigma, USA) was added to the YP medium.

The glucose content of the different calibration solutions (buffer, YP medium, YP medium + NaCl) was adjusted according to the different concentrations studied (1 – 150 mM, starting with 1 mM, followed by 5 mM and continued in steps of 10 mM up to 150 mM). A proper volume of a 150 mM glucose stock solution prepared in the according background matrix was added to the respective matrix to obtain 10 ml of the final calibration solution. In this way, the background of each medium was obtained while altering the sugar concentration.

1.2.3 Off-line glucose measurements in yeast fermentation samples

A yeast fed-batch fermentation was performed cultivating the classical laboratory yeast strain CENPK-113 7D in YPD medium. The fermentation was run in a 2.5 liter glass vessel, equipped with dissolved oxygen (DO), pH and temperature probe, the whole set up controlled by an Applikon ez controller. The process was started in a total volume of 2 liter and an initial glucose concentration of 150 mM. Controller and fermenter equipment were purchased from Applikon, The Netherlands. As broadly found in fermentation processes, glucose was used as the major carbon source by the microorganism. The fermenter was inoculated with 180 ml overnight culture, which was pre-grown in YPD medium using two 500 ml shake flasks (100 ml working volume each) at 30 °C, 180 rpm for 12-14 h. The fermentation process was run at 30 °C, a stirrer speed of 800 rpm and an aeration rate of 1 vvm. The pH was maintained at pH 6 using 2 M KOH and 2 M H₂SO₄. The dissolved oxygen tension (DOT) stayed above 30 % of saturation throughout the fermentation process, indicating that no oxygen limitation occurred during the cultivation. The first sensor was used for off-line glucose measurements in real-time. The process was followed over a 13 h period, during which glucose was fed three times. For the determination of glucose and biomass, samples were manually withdrawn every hour. Glucose was quantified with the biosensor in both, cell-free and a cell-containing samples, based on the calibration curve obtained in YPD medium. Hence, for every time point, a first sample was taken and immediately filtered via a 0.2 µm filter. A second sample, still containing the cells, was simultaneously pumped through the sensor. When the glucose measurement of this cell-containing sample was completed, the second, cell-free sample was measured via the sensor. Additionally, for each time point, a cell-free sample was stored in the fridge and quantified by HPLC after the fermentation process was ended. The glucose results from the cell-free samples, measured by the biosensor, were used to follow the consumption of glucose during the process. When the glucose concentration reached a value close to zero, 100 ml of a glucose feeding solution (500 g/l dissolved in purified water) was added via the septum by the help of a sterile syringe to prolong the exponential growth phase (addition after 7, 9 and 11 h). By doing so, the biosensor performance was challenged with respect to both, high cell and high glucose concentrations simultaneously.

The increase of biomass was followed by classical optical density (OD) and dry weight off-line measurements.

The OD was determined at 600 nm with the UV-1800 spectrophotometer from Shimadzu, Germany. Dry weight was determined by filtering 5 ml of each sample via a 0.2 µm filter-paper (Cellulose Nitrate Membrane Filters, Whatman, Germany) and washing it three times with 5 ml of purified water applying vacuum filtration. The filter cake, consisting of the washed biomass, was dried in the microwave for 15 min at 180 W. The weight of the filter-paper was determined before (clean filter paper) and after filtration and drying (filter paper with biomass) in order to obtain the dry weight of each sample. Dry-weight and OD measurements were performed in duplicates. Additionally the increase of biomass was followed by on-line backscatter data using the non-invasive Cell Growth Quantifier CGQBIOR from aquila biolabs (Baesweiler, Germany). The CGQBIOR was studied in detail in chapter 4.

HPLC was used as a reference method for glucose quantification in calibration and fermentation samples. The measurements were performed on the Ultimate 3000 Dionex HPLC system (Sunnyvale, USA) using an Aminex HPX 87 H column, 300 x 7.8 mm (BIORAD, Denmark) operated at 50 °C equipped with Refract Max 520 refractive index (RI) detector. The column was operated with 5mM H₂SO₄ in purified water as mobile phase and a constant flow rate of 0.6 ml/min. Samples were injected with a volume of 5 µl for analysis. Samples were filtered via a 0.2 µm filter and acidified (950 µl sample + 50 µl 5M H₂SO₄) prior to analysis. Acidification was required due to the ion exchange principle of the column used. Sample quantification was done with the software Chromeleon 6.8.

1.2.4 On-line glucose monitoring during a yeast fermentation

For operating a second biosensor in an on-line setup, it was connected to the fermenter via a recirculation loop. The flow was generated with the Ismatec Reglo ICC pump (Cole-Parmer, UK) circulating the fermentation broth via a sampling port to the sensor and back to the fermenter (using a separate inlet). The dead volume inside the tubing until the sensor was approximately 1 ml, applying a flow rate of 1ml / min. The flow rate was chosen according to the dead volume to ensure a reasonable exchange of the flow through cell with fresh fermentation broth (within this set up approximately every min). As suggested by the company, the maximal flow rate when operating the sensor should not be higher than 1 ml/min, which is typical for microfluidic devices. In order to ensure a reasonable exchange rate of the fermentation broth inside the flow-cell and to avoid oxygen limitation inside the tubing to the biosensor, the sensor should be close to the fermenter. Besides, in order to avoid cells sticking to the membrane inside the flow cell, operation at a higher flow rate may be desirable. The sensor used was non-sterile. However, a sterile version is available by the company. No contamination was observed during the fermentation as confirmed by microscopy of a mid- and end-fermentation sample. During this run, the fermenter was equipped with two sampling ports, one used for the recirculation loop, the other one used to withdraw samples manually every hour to validate the sensor data by HPLC measurements.

1.2.5 Dissolved oxygen conversion inside fermentation samples

The dissolved oxygen conversion rates were measured for the withdrawn fermentation samples in parallel with the biosensor characterization. The idea behind this experiment was to confirm the

competition for the oxygen consumption between the active cells present inside the sample and the enzymatic layer of the biosensors system. Thus, similar to the previously described set-up assembly and procedure in [22], dissolved oxygen (DO) was measured using OXR430 retractable needle-type fiber-optic oxygen minisensor (PyroScience GmbH, Aachen, Germany) connected to a FireStingO2 fiber - optic meter (PyroScience GmbH, Aachen, Germany) and controlled by Pyro Oxygen Logger software (PyroScience GmbH, Aachen, Germany). The time courses of oxygen profiles were registered at room temperature with constant stirring speed until the dissolved oxygen was depleted by the cells.

1.2.6 Storage conditions of the sensor

After usage the sensor was flushed with purified water for around 10 min (0.5 ml/min) in order to rinse out all forms of molecules potentially trapped inside the flow-through-cell and afterwards dried with compressed air. Luer connections of the inlet and outlet tubing were closed. The washed and dried sensor was stored inside a small bag in the fridge, together with a desiccant.

1.3 Results and Discussions

1.3.1 Investigations of sensor 1

A first sensor was studied with respect to the application of flow (0.2 ml/min) vs. non-flow, calibration in different media, application on fermentation samples and performance after usage and a 3 month storage period (long-term stability), see [Table 1.1](#).

1.3.1.1 Sensor calibration

The sensor was calibrated at room temperature using three different calibration media (buffer, YPD medium and YPD medium containing 9 g/L NaCl), preheated to 30 °C. They were analyzed with (0.2 ml/min) and without the application of flow. Calibration was done against glucose concentration values obtained by HPLC analysis of each calibration solution. As the calibration solutions were prepared manually, a small off-set compared to the target concentration value can be expected. Although the concept of mass transport fully supports the application of flow during the measurement with the biosensor, I wanted to have a look at the performance of the sensor with minimum technical effort, practically meaning nothing more than a syringe is needed to load the sensor with sample. [Figure 1.2 A-C](#) shows the calibration curves, or signal profiles, respectively, in all three media. All profiles are presented as current [nA] as a function of glucose concentration [mM] under non-flow conditions ([Figure 1.2 A](#)), flow conditions ([Figure 1.2 B](#)) and an overlay of both data sets ([Figure 1.2 C](#)) for simplified visual comparison. As expected, generally the profile followed a saturation curve. [Figure 1.2 D](#) shows the raw signal development as a function of time for each calibration point obtained in YPD medium, applying flow.

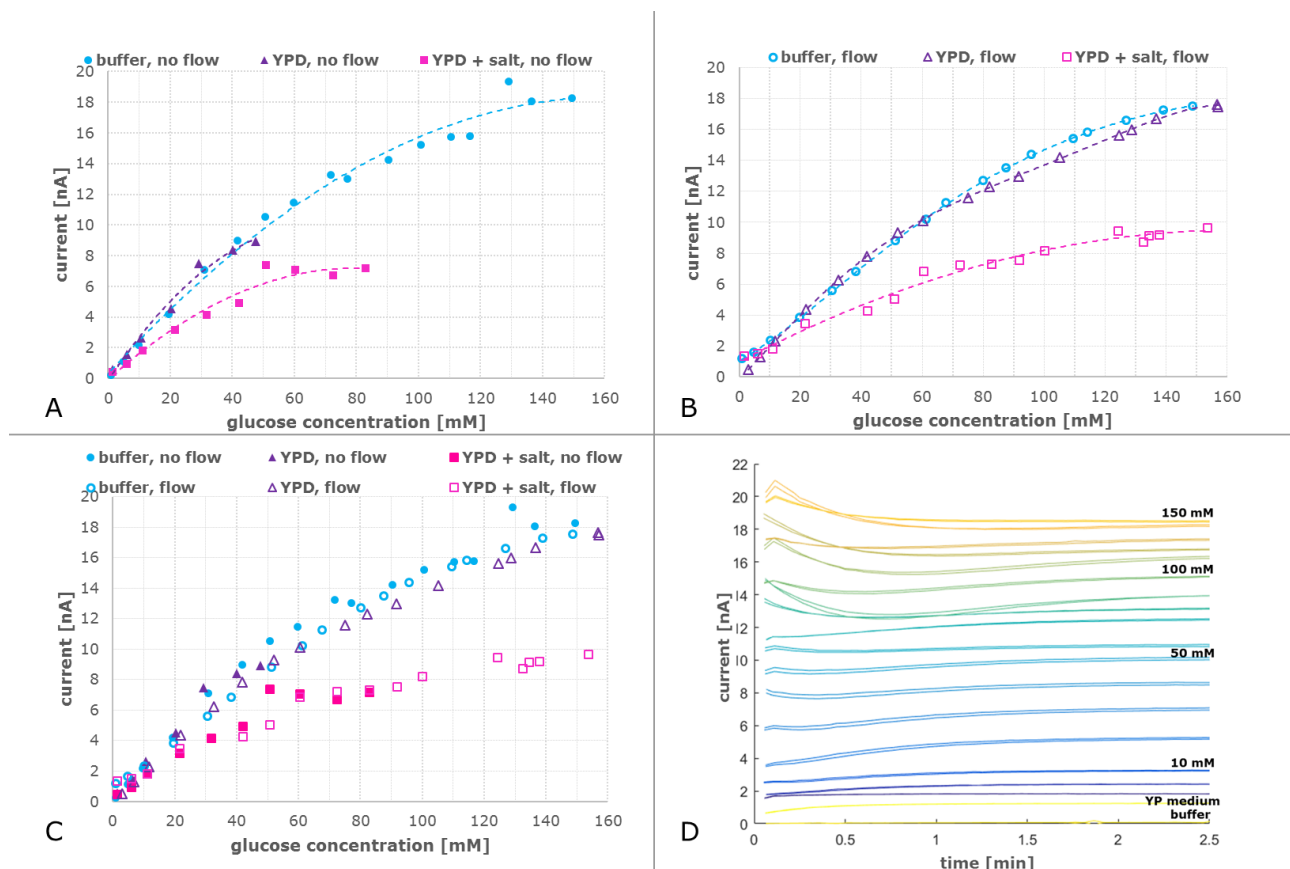


Figure 1.2: A–C) Current in nA (value obtained from the last 10 steady state values averaged, subtracting the respective blank current values from the glucose signal) as a function of glucose concentration in mM in the 3 different media investigated under non-flow (A) and flow (B) conditions. The dotted lines in A) and B) show the polynomial fit for easier visual comparison. C) Overlay of both, non-flow and flow data sets for simplified visual comparison. D) Raw signal (current in nA as a function of time in min) development for glucose concentrations obtained in YPD medium applying a flow rate of 0.2 ml/min. The lowest current curves (yellow) were obtained in buffer and YP medium as labelled, followed by a glucose concentration of 1, 5 and 10 mM. The steps between the indicated concentration values of 10, 50, 100 and 150 mM are 10 mM. Every calibration solution was measured in duplicate.

The signal development of each measurement can be found in the supplementary material (Figure S1.1 – S1.6). [Figure 1.3 A](#) and [B](#) show the average standard deviation and the average time until signal stabilization for the three different data sets.

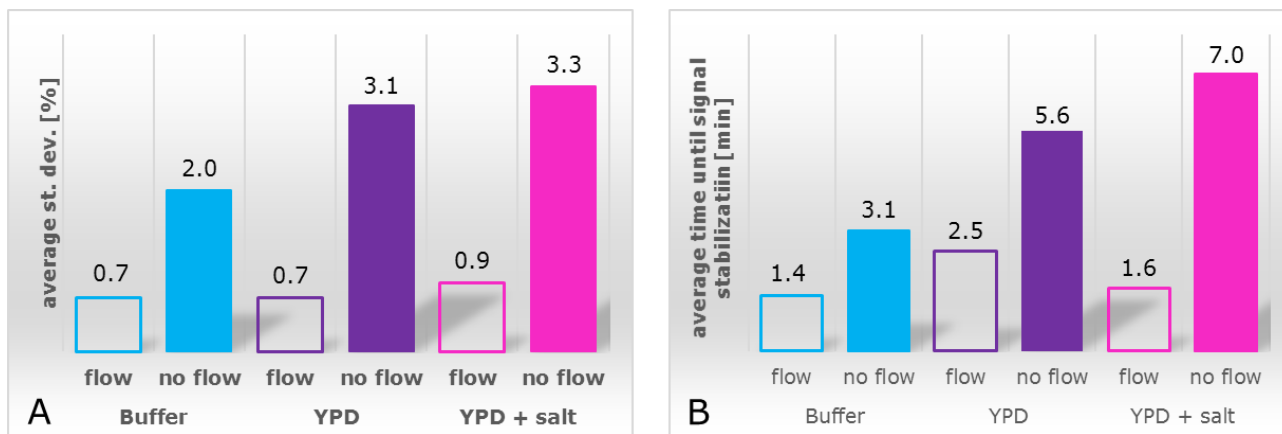


Figure 1.3: A) Average standard deviation of the final current in % in the three different data sets (buffer, YPD, YPD + salt). Average time until signal stabilization in min in the three different data sets. For Figure 1.3 A and B, the value per data point (averaged from a duplicate measurement) can be found in the supplementary material, Table S1.1 A and S1.1B.B).

As shown in Figure 1.2 A and B, the application of flow smoothed the signal profile while the overall trend in each medium stayed the same under flow and non-flow conditions (Figure 1.2C). This is confirmed by the average standard deviation (st. dev.) for each data set (Figure 1.3 A). In the absence of flow, the lowest st. dev. was found in buffer (2 %), followed by YPD medium (3.1 %) and YPD medium containing additional salt (3.3 %). Applying flow during the measurement reduced the st. dev. to < 1 % independent of the medium.

From Figure 1.2 A-C, it also becomes obvious that the signal profile obtained in YPD medium is very close to the signal profile obtained in the recommended acetate buffer, especially within the application range up to a glucose concentration of 60 mM. At glucose concentrations between 60mM – 150 mM, the profiles obtained in YPD medium and buffer deviate slightly. However, a glucose concentration of approx. 150 mM resulted in both cases in a current of approx. 18 nA. The addition of NaCl significantly lowered the overall signal profile (Figure 1.2 A-C) and a glucose concentration of 150 mM results in a current of 9 nA, half of the current value obtained with the same glucose concentration in YPD medium and buffer.

In the absence of flow, the average time until signal stabilization was clearly dependent on the medium composition (Figure 1.3 B). It was lowest in buffer (3.1 min) and highest in YPD including additional NaCl (7 min). Under flow conditions, the time until signal stabilization was more than 50 % reduced compared to non-flow conditions in buffer and YPD medium (1.4 and 2.5 min), and more than 75 % lower in YPD medium containing additional NaCl (1.6 min).

The enhanced sensor performance with flow can be explained by an increased mass transfer of glucose to the membrane due to a reduced thickness of the diffusion layer on top of the membrane. Besides, H^+ ions produced during the reaction with the enzyme (equation 1.3) are flushed out, thereby avoiding a local acidification, potentially decreasing the enzyme activity. Hence, flow was found to be the operation of choice when using the sensor. The application of flow yielded a st. deviation of less than 1 %, on average, independent of the medium under investigation (Figure 1.3 A). Measurements in the absence of flow led to less accurate and reproducible results compared to measurements under flow. However, with an average standard deviation of less than 5 % (Figure 1.3 A) they can still be considered as fairly reliable, particularly for glucose concentrations ranging up to 60 mM. NaCl addition was found unnecessary and even reduced the sensitivity of the sensor.

The reduced sensitivity in the presence of added NaCl might be explained by a chloride monolayer adsorbed onto the electrode surface, hindering the electrochemical kinetics. Monolayer coverage of chloride ions on platinum electrodes is described to occur at chloride concentrations of 100 mM [23]. Since in the present work, around 150 mM chloride ions were present and halide ions generally show a strong tendency to adsorb on platinum, it is likely that (monolayer) adsorption of chloride ions on platinum electrode caused resistance to the electrochemical reactions, thus reducing the sensor sensitivity.

The sensor is rated for glucose concentrations ranging between 0.05 mM and 60 mM. This also appeared to be the reliable linear range in which the current at a certain glucose concentration was found to be nearly independent of the medium used and the application of flow. (Figure 1.2 A- C). Flow was desirable as it certainly improved the reproducibility of the signal (below 1 % standard deviation with flow on average compared to an average st. dev. < 5 % without flow, Figure 1.3 A) and decreased significantly the average time until signal stabilization of the sensor (with flow less than 2 min compared to values of 3 – 7 min on average without flow, Figure 1.3 B).

Since this study was aiming at the application of the sensor on fermentation samples, the calibration curve was based on the signal profile obtained in YPD medium applying flow. The calibration curve was divided into two sections, approximated by linear regression. It is presented in Figure 1.6 D, as part of the section 1.3.3 Batch-to-batch variability and sensor stability. By segmentation of the calibration curve, the full glucose concentration range typically found in various fermentation processes (0 - 150 mM) could be covered without the necessity of sample dilution. The R^2 value for both sections was greater than 0.97, hence suggesting that glucose quantification could be performed reliable.

1.3.1.2 Off-line glucose measurements on yeast fermentation samples

Figure 1.4 demonstrates the progress of the yeast fermentation process over 13 h. The microbial growth was followed by off-line OD600 and dry weight measurements, as well as on-line by means of a backscatter cell (Figure 1.4 A). The glucose concentration in manually withdrawn fermentation samples was measured off-line by applying the biosensor to cell-free and cell-containing samples as well as via HPLC analysis after the fermentation (Figure 1.4 B).

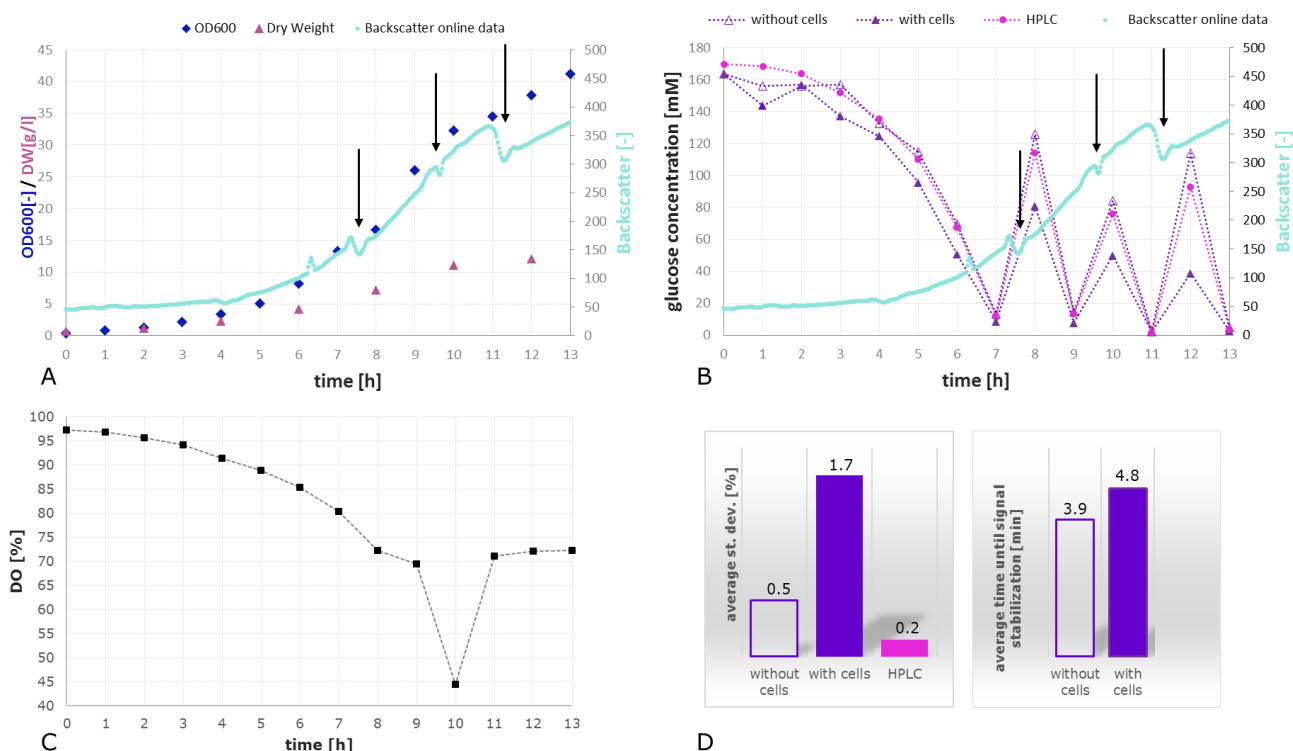


Figure 1.4: A) Biomass concentration as a function of time indicated by classical OD600 and dry weight measurements (off-line) as well as on-line data from a backscatter cell. In the backscatter data, the addition of glucose (indicated with black errors) can be seen in a small signal drop due to dilution of the fermenter content in the detection area. B) Glucose concentration measured off-line with and without cells by means of the biosensor and HPLC, respectively, as well as the biomass concentration in backscatter units as a function of time. C) Dissolved oxygen (DO) profile. The drop after 9 h is attributed to a disturbance in the air supply line. D) Left: Average standard deviation (st. dev.) of the different glucose measurements performed. Right: Average time until signal stabilization for glucose measurements by means of the biosensor. The values per data point can be found in the supplementary materials Table S1.2.

As shown in Figure 1.4 D, all glucose measurements performed with the biosensor were subject to a st. dev. of less than 2 %. On average, HPLC measurements showed a st. dev. of 0.2 %, biosensor measurements showed a st. dev. of 0.5 % in cell-free and 1.7 % in cell-containing samples. The increased st. dev. when measuring cell-containing samples by means of the biosensor can be expected. Cells generally increased the complexity of the samples, adding a solid phase, thus decreasing glucose diffusion to the enzyme layer. Besides, the cells kept consuming glucose during the measurement.

Furthermore, from Figure 1.4 B, it can be concluded that, firstly, glucose concentrations in the cell-free supernatant could be measured accurately by means of the biosensor based on the segmented calibration curve shown in Figure 1.6 B and are in line with the HPLC results. Secondly, glucose concentrations in cell-containing samples resulted in generally lower values compared to measurements on the cell-free supernatant. This was particularly the case for glucose concentrations above 20 mM. For glucose concentrations below 20 mM, no such an off-set was observed. For glucose concentrations above 20 mM (first 6 h of fermentation), there was an off-set (-10%) of biosensor measurements on cell-free samples compared to HPLC results. In the second half of the fermentation, subject to 3 times glucose spiking, the off-set for glucose concentrations exceeding

20 mM increased to ca. minus 40 % (comparing cell-free and cell-containing samples). This can be attributed to oxygen limitation occurring during the measurement with the biosensor. Both, the sensor and the cells competed for the present oxygen while the overall oxygen level in the fermentation broth was decreased to 70 % (Figure 1.4 A and C). Oxygen limitation generally hinders a full signal development with the biosensor and thus, the final current measured could not reach its actual steady state value. This could clearly be observed in a signal drop in the signals after 4, 5, 6, 8, 9 and 12 h (supplementary material, Figure S1.8). After 10 h, the signal development was not hindered. As can be seen in Figure 1.4 C, the DO dropped from 70 % to 45 % between 9 and 10 h, due to a disturbance in the air supply line. Since yeast is able to grow under oxygen limited conditions, the full signal development after 10 h might be explained by a metabolic shift towards anaerobic growth, meaning that the O₂ uptake by the organism was reduced. Thus, the remaining oxygen available for the sensor would be enough to enable a full signal development. Disregarding the exception after 10 h, Oxygen limitation was observed at glucose concentrations above 20 mM and cell concentrations above 5 g/l dry weight. For glucose concentrations below 20 mM, no oxygen limitation was observed within the fermentation reaching a max. cell dry weight concentration of 12 g/L and a minimum DO level of 70 % inside the fermenter.

As can be seen in Figure 1.4 D, the average time until signal stabilization was around 4 min for cell-free samples and around 5 min for cell-containing samples. The increased time until signal stabilization for cell-containing samples can be expected. It can be explained by both, decreased diffusivity of glucose to the enzyme layer in the presence of cells and oxygen competition between the sensor and the cells, decreasing the activity of the enzyme. However, a trained operator can obtain results in less than 5 min, counting from the time of sampling. The time- limiting step is the signal development until steady state.

1.3.2 Continuous on-line glucose measurements during a 10 h yeast fermentation

A second sensor was calibrated in buffer and YPD medium under flow conditions in order to investigate batch-to-batch variability between sensors. The calibration curve obtained in YPD medium was subsequently used for continuous glucose monitoring during another yeast fed- batch fermentation, conducted for 10 h. Additionally, (cell-containing) samples taken manually during the continuous glucose monitoring set-up were investigated regarding their oxygen content or, respectively, their oxygen consumption rate over time. This was done in order to confirm that the decreased glucose concentration obtained by means of the biosensor compared to HPLC results in cell-containing fermentation samples (sensor 1, Figure 1.4 B) was compromised by oxygen consumption of the cells. The data collected during the continuous fermentation set-up is presented in Figure 1.5 A-D. The data comprises the raw signal of the sensor (Figure 1.5 A), the converted glucose signal of the sensor (Figure 1.5 B), the OD600 and DO profile during the fermentation (Figure 1.5 C) and the oxygen consumption by cells measured in manual samples from the reactor (Figure 1.5 D).

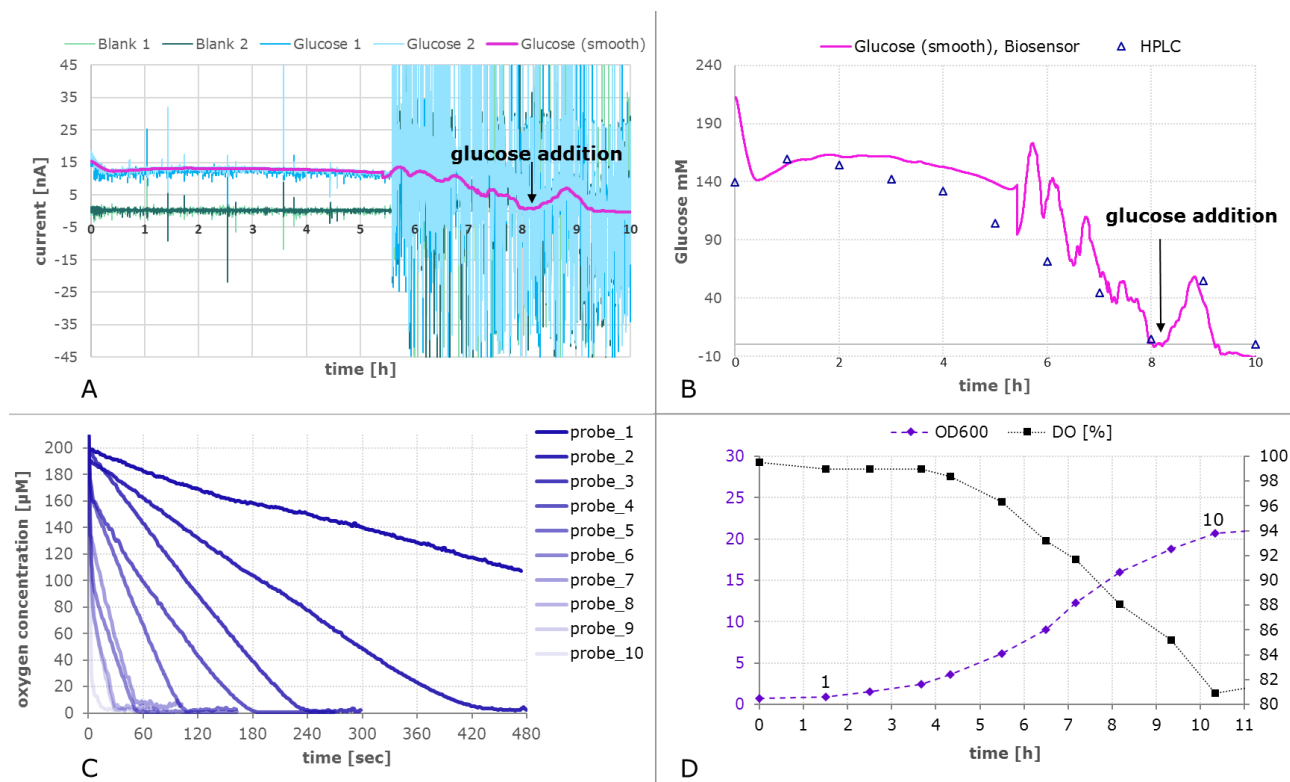


Figure 1.5: Data collected with sensor 2 during continuous glucose monitoring of a fed-batch yeast fermentation. A) The raw signal, current in nA as a function of fermentation time in min, as well as the smoothed average of the glucose signal obtained from a built-in Matlab function. B) Glucose concentration in mM as a function of fermentation time in min - Comparison of the continuous biosensor signal (on-line) and HPLC results (of manual off-line samples). The glucose concentration obtained with the biosensor was calculated based on the smoothed raw signal in A) and the calibration curve obtained in YPD (Figure 1.6 B). C) Oxygen concentration in μM as a function of time in sec in manually withdrawn fermentation samples (cell-containing). The samples are labelled with probe 1 to 10 relative to their acquisition time indicated in the OD600 profile, Figure 1.5 D. D) Yeast growth indicated as OD600 (violet rotated square, left y-axis) and dissolved oxygen tension (DO, black square, right y-axis) as a function of fermentation time in hours. The numbers above the OD600 data points indicate the time point of probes 1-10 in Figure 1.5 C.

Figure 1.5 A shows the raw signal of the sensor over time. The signal became especially noisy after approximately 5.5 h. This noise as well as current spikes observed in the signal before 5.5 h can be attributed to an air bubble stuck in the sensor-cell or air bubbles passing the sensor-cell, respectively. As no precaution regarding the sampling of air (inherently occurring in aerated fermentation broth) was taken, keeping the setup simple, air bubbles could be observed in the tubing of the recirculation loop and are considered to cause the noise observed. Small air bubbles passing the system seemed to be a minor disturbance (until 5.5 h). Contrarily, an air bubble stuck inside the electrochemical cell (after 5.5 h) evidently caused tremendous noise making it difficult to see the actual signal trend. The actual trend of the glucose signal could be recovered by an in house smoothing function run in Matlab R2016a applying a differential filter based on [24]. The blank signal was neglected during this run as found to be unreasonable high after smoothing and thus the glucose signal was obtained based on a calibration curve disregarding blank subtraction. Besides, peaks in the smoothed glucose signal except for the peak indicated resulting from glucose addition (after approximately 8 h) must be considered as artefacts resulting from the noise. If no filtering (smoothing) can be applied to the signal, the sampling of air can be avoided by adding e.g. a 20 μm stainless steel filter cap to the sampling port. Generally, the glucose signal trend was captured accurately and was in good

agreement with HPLC results (Figure 1.5 B). The signal drop in the very beginning can be attributed to the addition of the inoculum, diluting the glucose concentration present in the fermentation broth. Note that concentration levels and not total mass values are presented in Figure 1.5 B. Hence, volume changes inside the reactor due to evaporation, addition of acid and base for pH control and withdrawing of samples were not taken into account. Thus, the glucose signal seems to increase at the beginning of the fermentation (measured both by the biosensor and HPLC) but it must be considered as an artefact caused by volume changes. Interestingly, the glucose signal measured by the sensor is constantly a bit lower than the HPLC measurements showing the opposite in Figure 1.4 B. As a lower sensor signal compared to HPLC measurements can be expected, resulting from the presence of cells as described with respect to Figure 1.4 B, this might be due to the filtering function applied to the raw signal.

Electrochemical GOx based biosensors are subject to a minimum oxygen availability during operation. The sensor under study requires approximately 10 μM dissolved oxygen for the accurate quantification of 40 mM Glucose [5]. As described before (Figure 1.4 B), cell containing samples analyzed off-line run into oxygen limitations when the glucose concentration is higher than 20 mM and the cell concentration is higher than 5 g/l cell dry weight, or an OD600 value of approximately 7, respectively. According to Figure 1.5 D, an OD600 value of 7 was reached after ca. 6 h. The respective oxygen consumption is represented in Figure 1.5 C, probe 6. After 6 h, the dissolved oxygen concentration in a sample was approximately 95 μM in the beginning of the measurement. It was consumed in less than 1 min, while the signal stabilization time of the glucose biosensor was around 5 min (Figure 1.4 D). This confirmed that a full signal development was not possible in cell-containing samples exceeding a cell concentration of 5 g/l dry weight (or an OD value of 7, respectively) as cells depleted the present oxygen before the sensor signal could reach steady state.

1.3.3 Batch-to-batch variability and sensor stability

Sensor stability and batch-to-batch variability is always a subject of interest when using a biologically active recognition elements (here the enzyme GOx). With respect to this interest, both, the 1st and the 2nd sensor were calibrated before and after the use on fermentation samples. Besides, the 1st sensor was calibrated again after a storage period of 3 months. While off-line measurements of fermentation samples were conducted with the 1st sensor, storing the sensor in buffer in between samples, the 2nd sensor was used in a continuous glucose monitoring set-up circulating the fermentation broth with 1 ml/min over 10 hours through the sensor. The results of the calibration behavior over time are presented in Figure 1.6 A and 1.6 C. Figure 1.6 B and 1.6 D show the calibration curves applied for glucose determination in fermentation sample (off-line or on-line) for sensor 1 and 2, respectively.

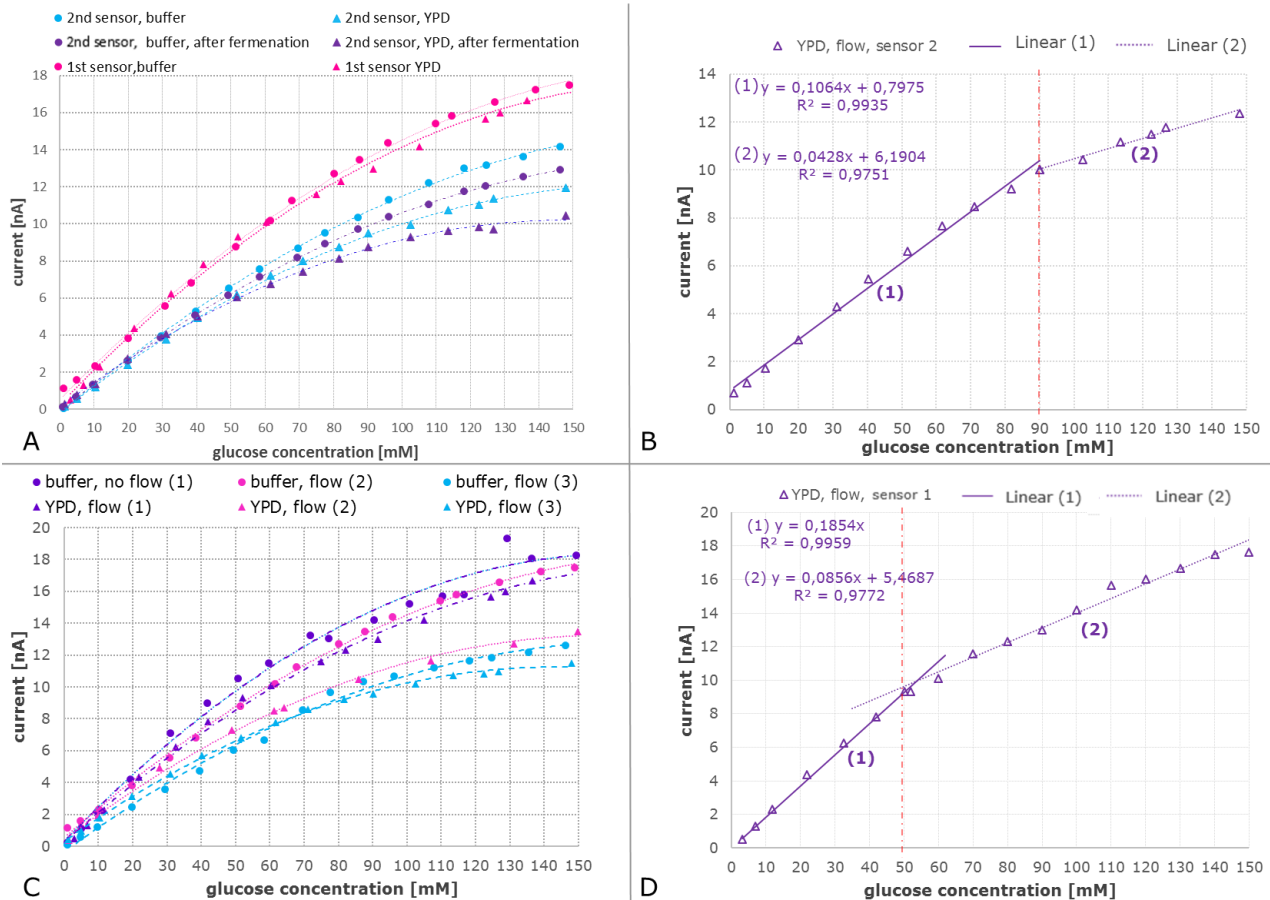


Figure 1.6: Different calibration profiles ranging from 1 -150 mM glucose concentration presenting the current [nA] as a function of glucose concentration measured by HPLC [mM]. A) Calibration profiles obtained with the 1st sensor in buffer and YPD medium before (1) and after the off-line use on fermentation samples (2) and after a storage period of three months after the first use (3). B) Calibration curve obtained in YPD medium under flow conditions (0.2 ml/min) used for continuous on-line glucose measurements during the first yeast fermentation. The curve (current in nA as a function of glucose concentration in mM) was divided into two linear calibration curves ranging from 1 to 50 and from 50 to 150 mM glucose (division point is marked with a red dashed line). The decision, which curve to apply, was based on the current value. C). Calibration profile of the second sensor in buffer and YPD medium under flow conditions (0.2 ml/min) before and after use in a continuous monitoring set up during the 2nd yeast fermentation. The data is compared to the calibration profile of the 1st sensor in buffer and YPD medium. D) Calibration curve obtained in YPD medium under flow conditions (0.2 ml/min) used for off-line glucose measurements during the first yeast fermentation. The curve (current in nA as a function of glucose concentration in mM) was divided into two linear calibration curves ranging from 1 to 90 and from 90 to 150 mM glucose (division point is marked with a red dashed line). Blank subtraction was neglected in this calibration curve. The decision, which curve to apply, was based on the current value.

Inspection of Figure 1.6 A - D reveals three main conclusions. Firstly, the correlation of current and glucose concentration was similar between the two sensors in buffer and YPD medium and sensor dependent (Figure 1.6 A and C). The first sensor showed an around 75 % higher sensitivity (0.1854 nA / mM glucose, Figure 1.6 D) than the second one (0.1064 nA / mM glucose, Figure 1.6 B) with respect to the low range section of the calibration curve. Both values are bisected in the high range calibration curve. Besides, the division between low and high range calibration curve was 50 mM for the first and 90 mM for the second sensor. However, the R^2 value for the low and high range calibration curve of both sensors was with 0.99 and 0.97 of reliable accuracy. The sensitivity towards glucose is higher for (low) glucose concentrations up to 60 mM (the recommended upper glucose limit from the company) but the sensor can be applied reliably for glucose concentrations up to

150 mM. The batch-to-batch variability is essentially linked to different enzyme activities and/ or quantities immobilized on the electrode.

Secondly, the sensitivity for glucose samples measured in buffer, especially for glucose concentrations in the higher range, was consistently a bit higher compared to measurements in YPD medium (Figure 1.6 A and C). This might be explained by a change of the diffusive properties of the membrane being decreased when large molecules like peptides are present as in the complex medium. Complex molecules might close pore structures and/ or attach to the surface thus decreasing the diffusivity inside the membrane or through the diffusion layer towards the membrane. Thirdly, the sensitivity decreased as a function of use and time (Figure 1.6 A and C). However, even after 10 h continuous use during a yeast fermentation, the sensor was active and could be reused as evidenced by the respective calibration curves performed before and after the fermentation. The loss in sensitivity can be linked to a loss in enzyme activity or in changes in the membrane characteristics. The operational stability stated by the company at 37 °C is more than 2 weeks in continuous operation at glucose concentrations up to of 25 mM.

1.4 Conclusions

Within this study, a commercial biosensor for glucose detection designed as a flow-through-cell, originally developed for medical applications was tested with respect to glucose detection in fermentation samples. Two sensors were studied. The first sensor was investigated regarding the calibration behavior in different media, real-time glucose measurements in fermentation samples conducted off-line and storage stability. The second sensor was tested with respect to batch- to- batch variability and applied on-line for continuous glucose monitoring over a 10 h yeast fed-batch fermentation. The biosensor was found to be a suitable device for real time glucose measurements conducted off-line and on-line.

For off-line glucose measurements, the average st. dev. was below 2 % for both, cell-free and cell- containing samples. Comparing biosensor and HPLC results, no off-set was observed for glucose measurements of the cell-free supernatant and in cell-containing samples with a glucose concentration below 20 mM. Especially low glucose concentrations (less than 20 mM) could be measured without significant off-set to HPLC results regardless of the cell concentration present (reaching a maximum of 12 g/l cell dry weight). Generally, for cell-containing samples and glucose concentrations above 20 mM, the measurement was compromised by oxygen limitation hindering a full signal development (when performed off-line). For the respective samples, an off-set to the reference value of minus 10 – 40 % was observed depending on cell and glucose concentration present. Note that, the combination of high glucose and high cell concentration must be considered as an exception when talking about fermentation processes. In this study, they were induced by glucose spiking during the fermentation process to challenge the sensor particularly with respect to high glucose and cell concentrations. The off-set within the initial batch phase, comparing cell-free and cell-containing samples, was on average 10 %. Normally, high glucose concentration are found in the beginning of batch processes when the cell concentration is low. For economical and metabolic reasons, feeding strategies are ideally designed to keep the glucose concentration in the broth as low as possible, feeding only as much glucose as is immediately consumed by the microorganisms. In such a case, the sensor can be considered as a valuable tool for monitoring and control of the glucose level. For continuous on-line glucose measurement, the measurement is expected not to be

compromised by oxygen limitations until very high cell densities, when the sensor is positioned close to the fermenter outlet. When used off-line, a trained operator can obtain results in less than 5 min, counting from the time of sampling. The time-limiting step is the signal development until steady state of the sensor signal is obtained.

The second sensor was successfully applied as a continuous glucose monitoring tool during a 10 h yeast fed-batch fermentation. Although a high level of noise occurred after approximately 5.5 h suspected to be caused by an air bubble stuck in the sensor, the signal was recovered applying a differential filter. The smoothened signal showed an off-set of about 10 %, compared to HPLC results. The trend of the glucose profile obtained by HPLC results was captured accurately with the sensor. When positioned close to the fermenter, on-line measurements seem to be desirable as oxygen limitations are avoided by supplying constantly fresh, aerated broth during the measurement. As expected, the sensitivity of the sensor was found to be dependent on storage time and usage whereas both sensors were reusable after the fermentation samples, and even after a subsequent storage period over 3 months (storage only tested with the first sensor). Neither of the sensors reached the end of their life time within this study. The sensitivity loss over time is linked to a loss in enzyme activity and changes in the membrane characteristics. Each sensor must be calibrated individually before use in a representative sample matrix as the calibration profile was found to be medium and sensor dependent.

The sensor was found to be a small, affordable, easy-to-use, and reliable tool facilitating both, continuous on-line and fast off-line measurements of glucose during a yeast fermentation process. I consider the sensor as especially valuable for the early stages of fermentation development, when microbial kinetics are still unknown. It will significantly reduce the time spent on process optimization. Applying the sensor, a first insight into the glucose consumption can be gained right away, allowing for optimization with respect to glucose feeding 'on the first go'. The confident use of applied research solutions starts in every laboratory before it will gradually find the entry to industrial application and processing. Only by using available technology in our own laboratories, we can build confidence and knowledge on their application and eventually tailor them towards our needs. With this study, I want to encourage people from relevant areas to gain confidence in considering and using commercially available biosensor solutions as a monitoring tool.

1.5 Supplementary Material

1.5.1 Figures

The following Figures S1.1 – S1.8 show the raw signal development, current in nA as a function of time in min, of the measurements performed with the sensor 1. Important here is the final current value in steady state as well as the time until signal stabilization. In each figure, the current of the two glucose electrodes (upper curves) and the current of the respective two blank electrodes (two lower curves) are presented. In most of the cases, the duplicate does not separate visually.

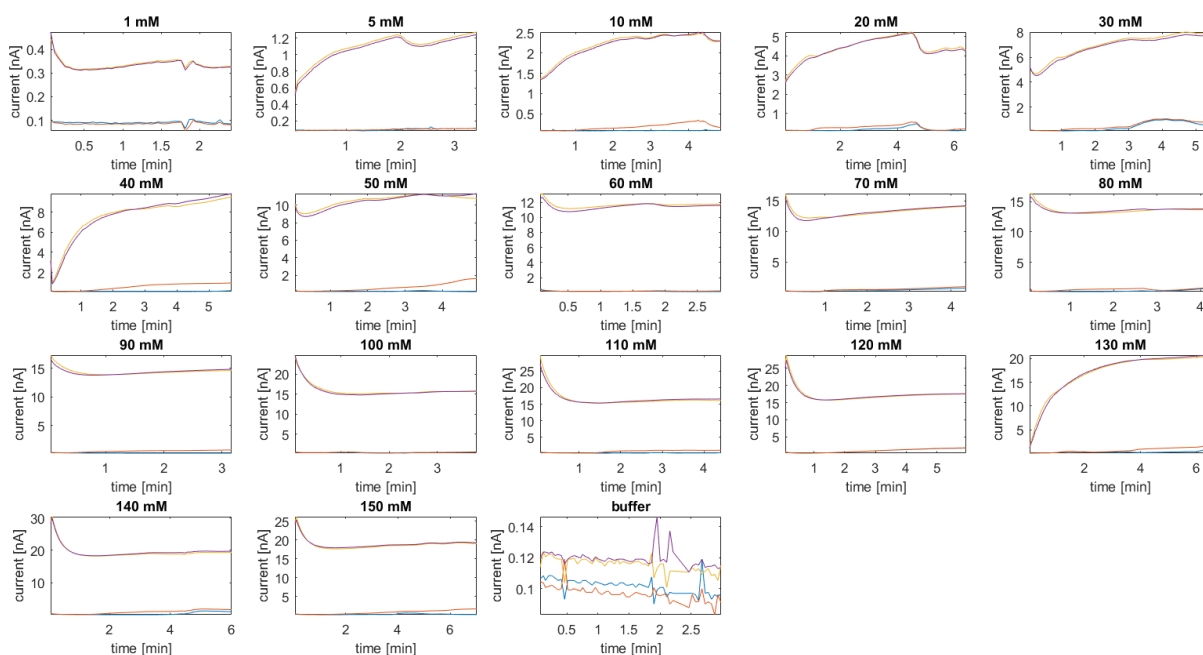


Figure S1.1: Sensor 1 - Raw signal development of measurements performed in buffer, no flow.

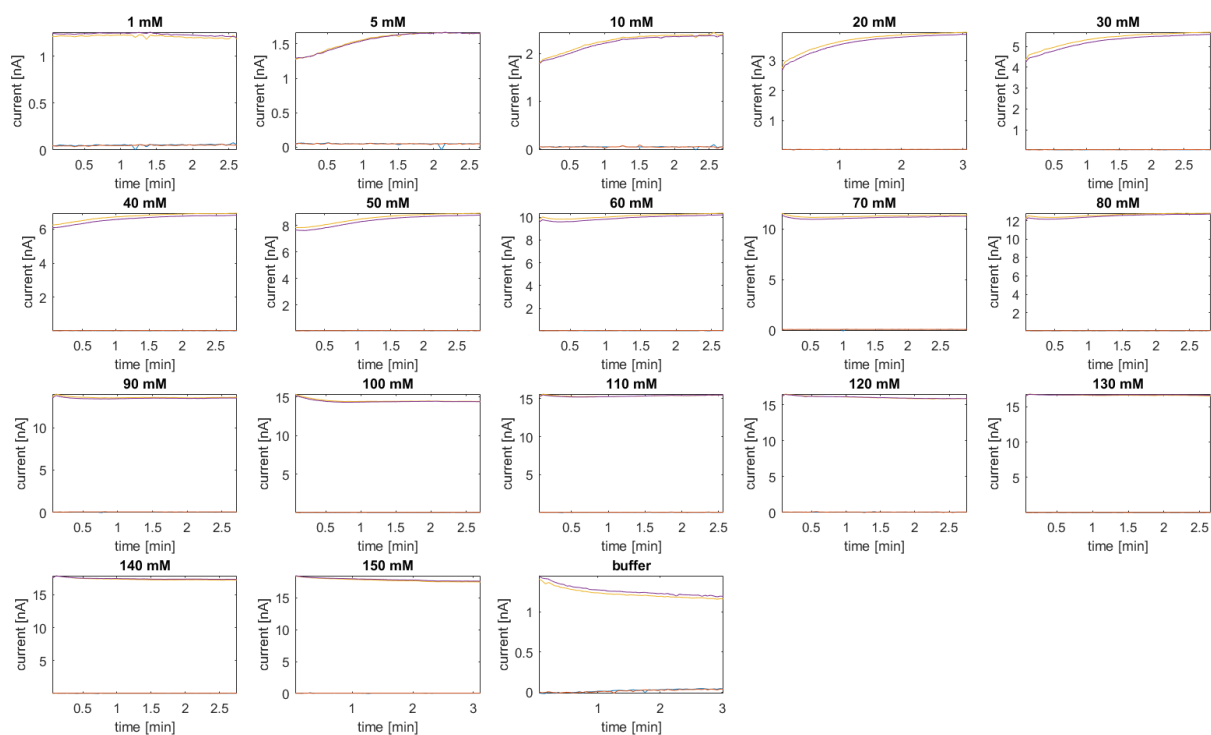


Figure S1.2: Sensor 1 - Raw signal development of measurements performed in buffer, applying a flow rate of 0.2 ml/min.

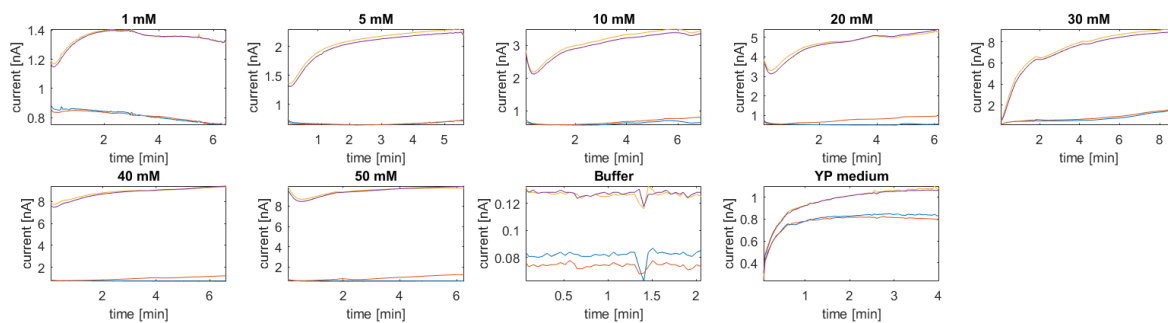


Figure S1.3: Sensor 1 - Raw signal development of measurements performed in YPD medium, no flow.

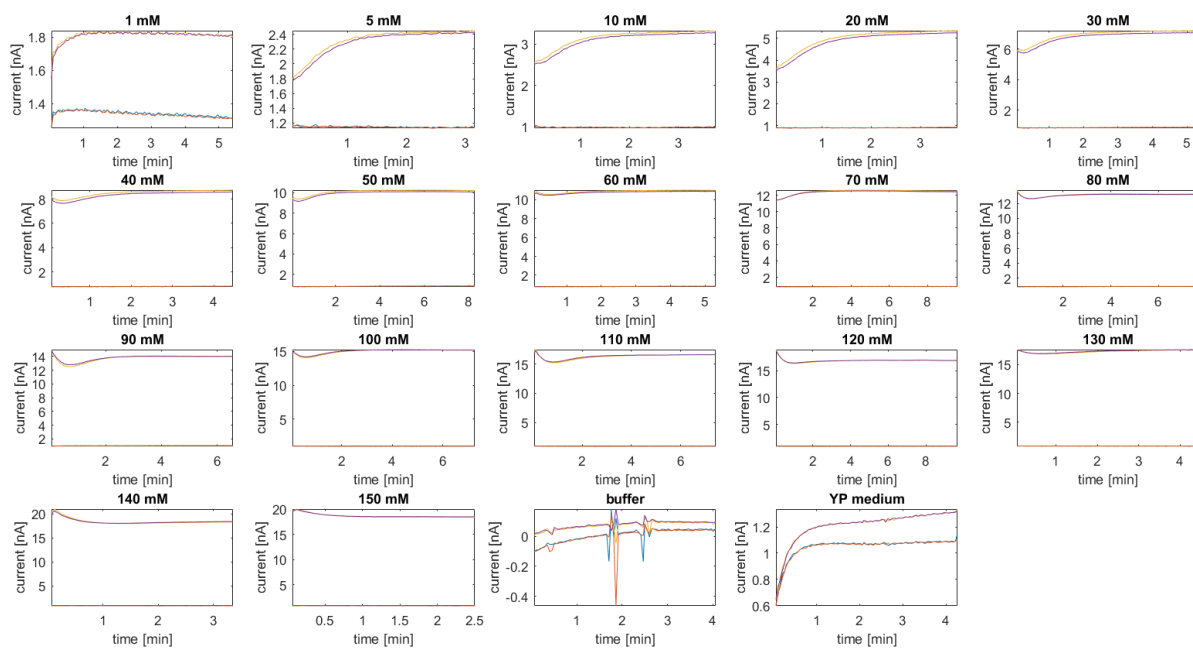


Figure S1.4: sensor 1 - Raw signal development of measurements performed in YPD medium, applying a flow rate of 0.2 ml/min.

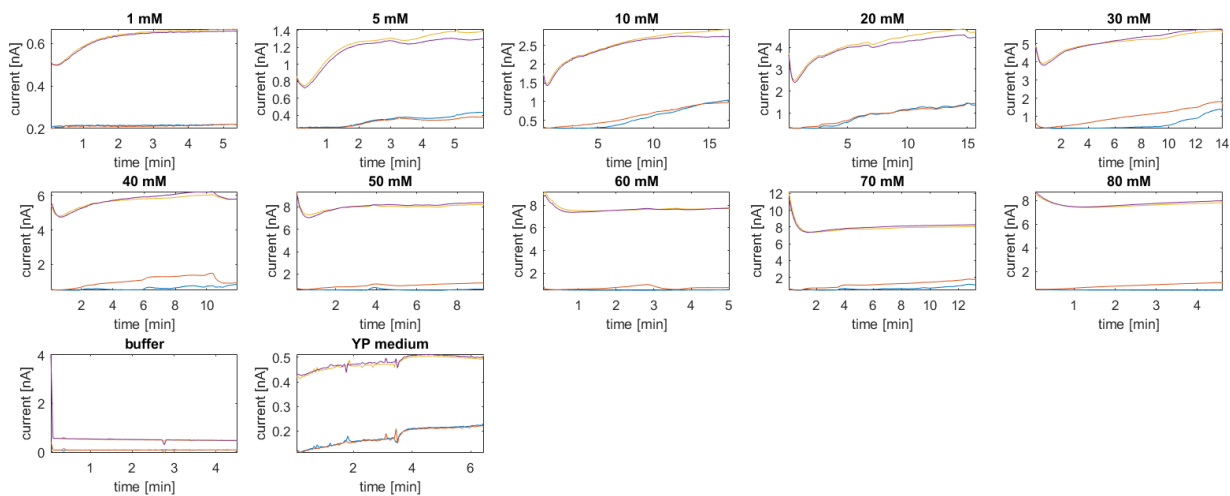


Figure S1.5: Sensor 1 - Raw signal development of measurements performed in YPD medium plus 9 g/L additional NaCl, no flow. The investigation was stopped after the sample of 80 mM glucose. Since the final current value did not change significantly after a sample concentration of 60 mM, it was assumed that the profile current as a function of glucose concentration had reached the plateau.

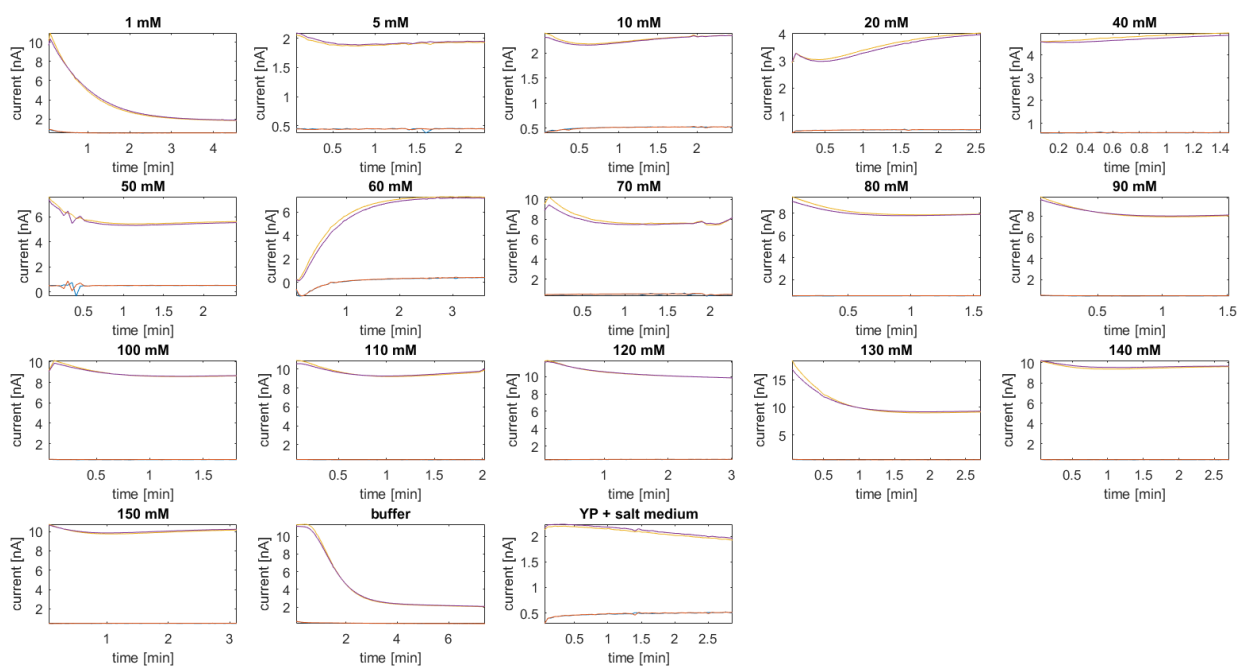


Figure S1.6: Sensor 1 - Raw signal development of measurements performed in YPD medium plus 9 g/L additional NaCl, applying a flow rate of 0.2 ml/min.

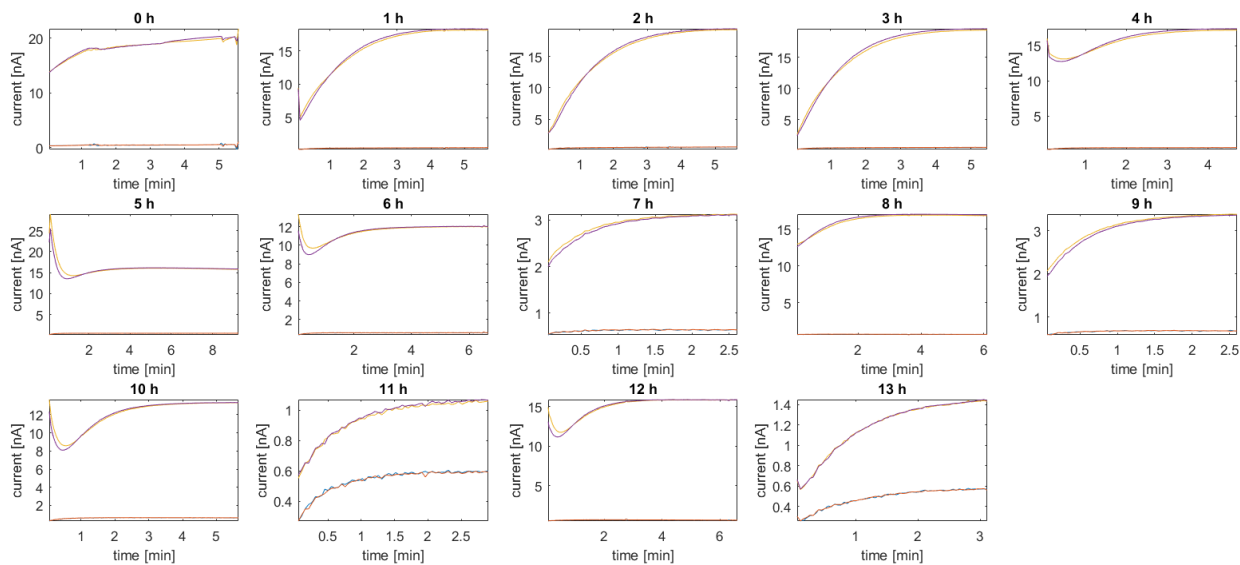


Figure S1.7: Sensor 1 - Cell-free fermentation samples, applying a flow of 0.2 ml/min.

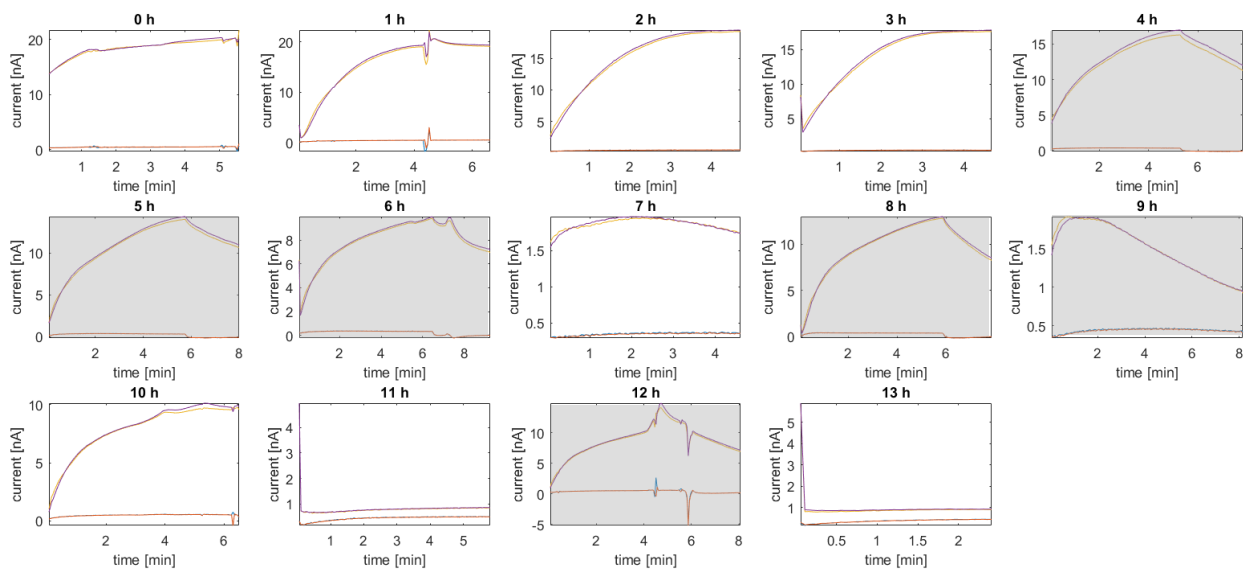


Figure S1.8: Sensor 1 - Cell-containing fermentation samples, applying a flow of 0.2 ml/min. The signals presented in the graphs with grey background are subject to oxygen limitations. The time until steady state was estimated.

1.5.2 Tables

Table S1.1A: Standard deviation of measurements performed with the biosensor in the three media investigated applying non-flow and flow conditions (0.2 ml/min). The average standard deviation (st. dev.) value of each data set is presented in the main text in Figure 1.3 A. (*indicates that this data refers to calibration data under flow conditions after the use of the sensor on fermentation samples. Otherwise, the data refers to the initial calibration curve obtained under flow conditions before the use of the sensor on fermentation samples.)

Table S1.1A: Standard deviation of measurements performed with the biosensor in the three media investigated applying non-flow and flow conditions (0.2 ml/min). The average st. dev. value of each data set is presented in the main text as Figure1.3A. (* indicates that this data refers to calibration data under flow conditions after the use of the sensor on fermentation samples. Otherwise, the data refers to the initial calibration curve before the use of the sensor on fermentation samples.)

theoretical glucose concentration [mM]	Buffer		YPD		YPD + salt	
	flow*	no flow	flow	no flow	flow	no flow
1	0.6%	0.7%	0.6%	1.6%	2.3%	0.4%
5	0.9%	1.8%	0.9%	2.0%	0.8%	2.8%
10	1.3%	3.6%	1.3%	5.0%	0.3%	4.2%
20	1.6%	2.8%	1.6%	3.9%	1.5%	4.3%
30	1.5%	4.0%	1.5%	3.0%	-	3.9%
40	1.4%	3.6%	1.4%	2.4%	2.0%	4.0%
50	1.1%	4.0%	1.1%	4.2%	1.5%	2.1%
60	0.9%	1.6%	0.9%	-	1.2%	3.7%
70	0.6%	0.7%	0.6%	-	0.1%	4.9%
80	0.5%	0.1%	0.5%	-	0.6%	2.7%
90	0.3%	1.7%	0.3%	-	0.6%	-
100	0.2%	0.5%	0.2%	-	0.1%	-
110	0.1%	1.8%	0.1%	-	0.6%	-
120	0.2%	0.1%	0.2%	-	0.1%	-
130	0.1%	1.9%	0.1%	-	1.4%	-
140	0.3%	1.0%	0.3%	-	0.4%	-
150	0.2%	4.7%	0.2%	-	0.5%	-
average st. dev. [%]	0.7%	2.0%	0.7%	3.1%	0.9%	3.3%

Table S1.1B: Time until signal stabilization of measurements performed with the biosensor in the three media investigated applying non-flow and flow conditions (0.2 ml/min). The time is estimated based on the respective figure S1.1-S1.6. The average time until signal stabilization of each data set is presented in the main text as Figure 1.3B.

	Buffer		YPD		YPD + salt	
theoretical glucose concentration [mM]	flow	no flow	flow	no flow	flow	no flow
1	0.5	1.5	2	4	4	4
5	2	3	2	5	1.5	4.5
10	2	4	2	6	2	13
20	2.5	5	2	5	2.5	15
30	2.5	5	3	8	-	14
40	2	5	3	6	1	10
50	2.5	4	3	5	1	3
60	2	2	3	-	3	2
70	1	3.5	3	-	1	3
80	2	2	3	-	1	1.5
90	0.5	2	3	-	1	-
100	1	2	3	-	1	-
110	0.5	2	3	-	1	-
120	0.5	2	3	-	2	-
130	0.5	5	2	-	1.5	-
140	0.5	2	1.5	-	1	-
150	1	2	1.5	-	1	-
average time until signal stabilization	1.4	3.1	2.5	5.6	1.6	7.0

Table S1.2: Time until signal stabilization and standard deviation (st. dev.) of measurements performed with the biosensor on fermentation samples applying a flow rate of 0.2 ml/min). The time is estimated based on the respective figure S7 and S8. The average values are presented in the main text as Figure 1.4 D. The grey marked fields are estimated times as the signal could not reach steady state due to oxygen limitation (Figure S1.8).

time	st. dev. [%]			time until signal stabilization [min]	
	cell free sample	cell sample	HPLC results	cell free sample	cell sample
0	0.2%	0.2%	0.0%	5	5
1	0.6%	0.9%	0.3%	5	5
2	0.5%	0.9%	0.1%	5	4
3	0.8%	0.8%	0.1%	5	4
4	0.8%	2.8%	0.2%	4	8
5	0.3%	1.6%	0.1%	4	8
6	0.1%	1.3%	0.1%	4	8
7	0.4%	1.4%	0.0%	2.5	2
8	0.4%	1.0%	0.2%	4	8
9	0.6%	1.3%	0.0%	2.5	2
10	0.1%	2.0%	0.0%	4	6
11	2.1%	3.8%	0.6%	3	1
12	0.0%	2.4%	0.2%	4	6
13	0.5%	3.8%	0.4%	3	0.5
average value	0.5%	1.7%	0.2%	3.9	4.8

1.6 References

- [1] C. Yan, F. Dong, B. Chun-yuan, Z. Si-rong, and S. Jian-guo, "Recent Progress of Commercially Available Biosensors in China and Their Applications in Fermentation Processes," *J. Northeast Agric. Univ. (English Ed.)*, vol. 21, no. 4, pp. 73–85, Dec. 2014.
- [2] FDA, Guidance for Industry, PAT — A Framework for Innovative Pharmaceutical Development, Manufacturing, and Quality Assurance, September 2004
- [3] M. Pohlscheidt, S. Charaniya, C. Bork, M. Jenzsch, T. L. Noetzel, and A. Luebbrecht, "Bioprocess and fermentation monitoring," in *Upstream Industrial Biotechnology: Equipment, Process Design, Sensing, Control, and cGMP Operations*, First., M. C. Flickinger, Ed. © 2013 JohnWiley & Sons, Inc., 2013, pp. 1469–1492.
- [4] "Encyclopedia of industrial biotechnology: bioprocess, bioseparation, and cell technology," *Choice Rev. Online*, vol. 48, no. 03, pp. 48-1218-48-1218, Nov. 2010.
- [5] I. Moser and G. Jobst, "Pre-calibrated biosensors for single-use applications," *Chemie-Ingenieur-Technik*, vol. 85, no. 1–2, pp. 172–178, 2013.
- [6] G. Jobst *et al.*, "Thin-Film Microbiosensors for Glucose-Lactate Monitoring," *Anal. Chem.*, vol. 68, no. 18, pp. 3173–3179, 1996.
- [7] L. C. Clark and C. Lyons, "ELECTRODE SYSTEMS FOR CONTINUOUS MONITORING IN CARDIOVASCULAR SURGERY," *Ann. N. Y. Acad. Sci.*, vol. 102, no. 1, pp. 29–45, Dec. 2006.
- [8] D. W. Hwang, S. Lee, M. Seo, and T. D. Chung, "Recent advances in electrochemical non-enzymatic glucose sensors – A review," *Anal. Chim. Acta*, vol. 1033, pp. 1–34, 2018.
- [9] K. Tian, M. Prestgard, and A. Tiwari, "A review of recent advances in nonenzymatic glucose sensors," *Mater. Sci. Eng. C*, vol. 41, no. April, pp. 100–118, 2014.
- [10] J.-M. Kauffmann and M. Pravda, "The Electrochemical Biosensor Era," *Bioforum Int.*, vol. 2, no. January 1998, pp. 22–25, 1998.
- [11] S. L. Brooks, R. E. Ashby, A. P. F. Turner, M. R. Calder, and D. J. Clarke, "Development of an On-line Glucose Sensor for Fermentation Monitoring," *Biosensors*, vol. 3, no. 1, pp. 45–56, 1987.
- [12] M. R. Phelps, J. B. Hobbs, and D. G. Kilburn, "An Autoclavable Glucose Biosensor for Micro b i a I Ferment at i o n Monitoring and Control," vol. 46, pp. 514–524, 1995.
- [13] E. S. P. B. V and B. Grundig, "Analytica Chimica Acta, 222 .(1989) 75-81," pp. 75–81, 1989.
- [14] J. Rishpon, Y. Zibenberg, R. Tor, A. Freeman, and T. R. Vessel, "Communications to the Editor In Situ Glucose Monitoring in Fermentation," vol. 35, pp. 103–107, 1990.
- [15] S. F. White, I. E. Tothill, J. D. Newman, and A. P. F. Turner, "Development of a mass-producible glucose biosensor and flow-injection analysis system suitable for on-line monitoring during fermentations," *Anal. Chim. Acta*, vol. 321, no. 2–3, pp. 165–172, Mar. 1996.
- [16] E. B. Bahadir and M. K. Sezgintürk, "Applications of commercial biosensors in clinical, food, environmental, and biothreat/biowarfare analyses," *Anal. Biochem.*, vol. 478, pp. 107–120, 2015.
- [17] P. Mehrotra, "Biosensors and their applications - A review," *J. Oral Biol. Craniofacial Res.*, vol.

6, no. 2, pp. 153–159, 2016.

- [18] J. Shi, D. Feng, and Y. Li, "Biosensors in Fermentation Applications," in *Fermentation Processes*, vol. 6, A. F. Jozala, Ed. InTech, 2017, p. 310.
- [19] S. S. Kamanin, V. A. Arlyapov, A. V. Machulin, V. A. Alferov, and A. N. Reshetilov, "Biosensors Based on Modified Screen-Printed Enzyme Electrodes for Monitoring of Fermentation Processes," vol. 88, no. 3, 2015.
- [20] H. Kalso, B. Gonz, S. Ma, R. Ludwig, and F. Bolado, "Dual Biosensor for Simultaneous Monitoring of Lactate and Glucose Based on Thin-layer Flow Cell Screen-printed Electrode," pp. 87–92, 2017.
- [21] C. Ampelli *et al.*, "Monitoring of glucose in fermentation processes by using Au / TiO₂ composites as novel modified electrodes," *J. Appl. Electrochem.*, vol. 45, no. 9, pp. 943–951, 2015.
- [22] D. Semenova *et al.*, "Mechanistic modeling of cyclic voltammetry: A helpful tool for understanding biosensor principles and supporting design optimization," *Sensors Actuators B Chem.*, vol. 259, pp. 945–955, Apr. 2018.
- [23] R. S. Patil, V. A. Juvekar, and V. M. Naik, "Oxidation of chloride ion on platinum electrode: Dynamics of electrode passivation and its effect on oxidation kinetics," *Ind. Eng. Chem. Res.*, vol. 50, no. 23, pp. 12946–12959, 2011.
- [24] P. H. C. Eilers, "Baseline Correction with Asymmetric Least Squares Smoothing," *Anal. Chem.*, vol. 75, no. 14, pp. 3631–3636, 2005.

Chapter 2

Monitoring of several fermentation parameters applying infrared (IR) spectroscopy and partial-least-squares (PLS) modelling

Preface

The subject of this chapter is the monitoring of a fermentation process using the combined approach of IR spectroscopy and chemometrics – more specifically Partial-Least-Squares (PLS) – modelling applied to spectroscopic data. It is exemplarily employed on a lab-scale yeast fermentation. The use of the novel nonlinear infrared spectroscopy (NLIR) technology as spectroscopic methodology, the investigation of indirect predictions as well as on-line spectral acquisition are the highlights and challenges within this chapter. The monitoring of the routine yeast performance parameters such as glucose (substrate), ethanol, glycerol and acetic acid (metabolites and products in relevant industrial production processes) is under study. Apart from this, the monitoring of ammonium and phosphate was considered due to a previous case study and submitted manuscript '*Fast measurement of phosphates and ammonium in fermentation-like media: feasibility study*', provided in appendix 1. Ammonium and phosphate are important nutrients in enzyme production processes and furthermore pose challenges to the wastewater treatment, as outlined in the manuscript. However, they are not of particular relevance for the studied yeast fermentation process, performed in a complex medium. Nonetheless, they were considered to continue the previous work and were shown to be useful for the evaluation of indirect predictions challenging this particular case study. The chapter leads the reader through all relevant aspects regarding the topic of infrared spectroscopy (IR) spectroscopy and PLS modelling. Theoretical concepts and considerations are linked to the respective practical issues and not presented in a separate section but 'where they make sense'. The introduction guides the reader towards this field while the most important considerations are presented as part of the main text, comprising material and methods, result and discussions and the final conclusions. Written in the form of an extended manuscript to be published in a relevant outline, I consider this chapter as a practical guide to the topic, for all readers starting on this highly interdisciplinary field.

The chapter was realized in close collaboration with The University of Copenhagen, Department of Food Science. In particular, I would like to express my great gratitude to Marta Bevilacqua, who performed the chemometric analysis as such and who helped to make this chapter possible to the extent presented.

2.1 Introduction

Within the PAT framework, the interaction of molecules with light has been widely used for their detection and quantification as demonstrated in [Figure 2.1](#). Due to their fast, high-throughput, informative and non-destructive nature, vibrational spectroscopic techniques, especially near (4000

– 13000 cm^{-1}) and mid (200 – 4000 cm^{-1}) infrared spectroscopy (NIRS and MIRS) [1]–[3], combined with multivariate data analysis [4]–[6] have been shown to be highly powerful for process understanding and control. In both cases, the infrared region of the electromagnetic spectrum, is employed for causing absorption bands characteristic for the common functional groups in organic molecules (alkyl, alcohol, carboxyl, carbonyl, amide, phosphate) in the near, or respectively, the mid IR region. Potentially, several components can be quantified within minutes (seconds) based on a single IR spectrum. Both, NIR and MIR spectroscopic techniques can be applied to the fermenter by fiber optical probes. Optical fibers used to transmit NIR light have a low self-absorbance, are considerably less expensive and can be much longer than MIR fibers. The latter, contrarily, have a high self-absorbance and hence their length is typically limited to two meters. Even though Attenuated Total Reflectance (ATR) fibers used to transmit MIR light have become more and more robust, they are still the most fragile part of the whole set-up and if not taken care can be damaged easily [7]. Besides, a MIR spectrometer typically involves moving parts in the optical core [8] and therefore is more sensitive to mechanical vibrations than a NIR spectrometer. It might be due to the more robust and flexible, yet less expensive design of a NIR spectroscopic set-up that NIR applications are more established in a production process environment than MIR spectroscopy. Another strong point of NIRS is the possibility to measure biomass and related parameters such as cell viability directly. This is possible due to the light scattering effects of cells and a significant change in the cell morphology (cell texture and size) when changing their physiological state [7]. Due to the low penetration depth in (ATR) MIRS, typically in the range of 0.5 – 2 μm [9], [10], it is not possible to measure the biomass concentration directly. Models based on MIRS predicting the cell concentration are based on indirect correlation with *e.g.* the glucose concentration and thus are very sensitive to process variation and glucose feeding.

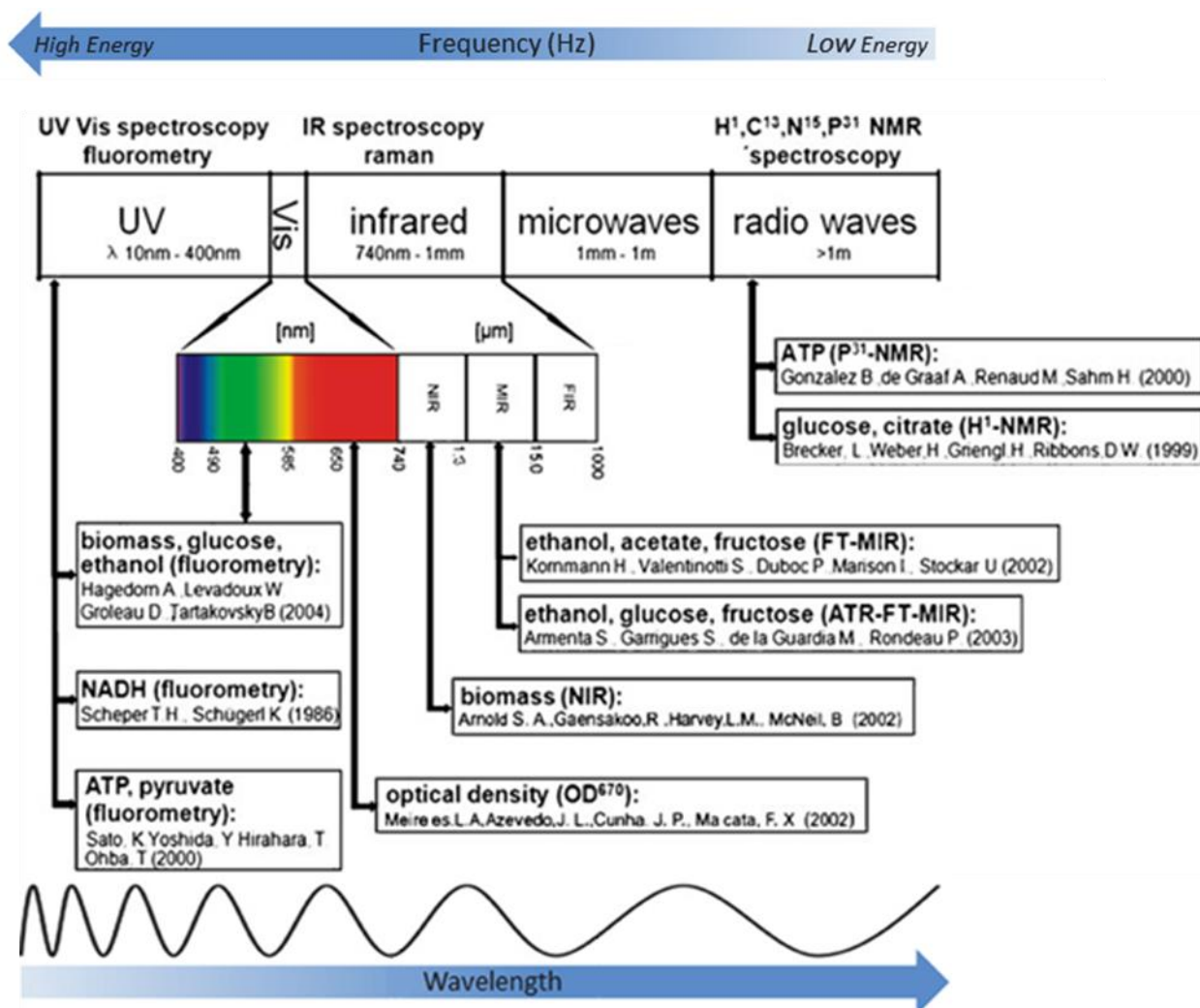


Figure 2.1: The electromagnetic spectrum and its spectroscopic application for biomolecule detection and quantification. Modified from [11].

However, MIR spectroscopy clearly scores higher than NIRS when it comes to selectivity, as nicely demonstrated in [12]. In most cases, fundamental molecular vibrations are excited by the MIR region whereas the overtones and vibration combinations are excited by the NIR region. Fundamental vibrations cause a bigger change in energy than the excitation of overtones and thus, MIR absorption yields more distinctive bands than NIRS, especially in the so-called fingerprint region from 800 - 1800 cm^{-1} . In the fingerprint region, MIR responds to all the functional moieties / groups constituting organic components. Thus, MIR produces exquisite molecular selectivity and its greater resolution enables the prediction of a single analyte at significantly lower concentration levels than with NIRS, especially in a complex culture medium as the fermentation broth. Besides, the main spectral variation in NIRS is due to the change in water content. The characteristic absorption bands of individual molecular bonds are strongly masked by water absorption bands. Thus, NIR is more likely to cause indirect calibration models. Indirect calibration models are not desirable for quantitative analysis as the prediction strongly depends on the process conditions, and the process dynamics must be highly conserved for the application of indirect models. Consequently, MIRS and in this field especially Fourier transform infrared spectroscopy (FTIR) have become increasingly interesting for

the monitoring of bioprocesses. Also MIRS is challenged by strong water absorption and thus cannot be applied directly to the sample. This has been solved by the introduction of ATR technology, which is considered as the fastest MIR technology with the highest level of reproducibility. Here, a diamond ATR is considered as the 'gold standard' [10], [12]. As mentioned before, current challenges limiting the practical application of the FT-IR technology to the academic environment are the high cost, the presence of fragile and sensitive moving parts, intrinsic thermal noise and the limited length of ATR fibers for remote sampling. These challenges are addressed by recent technical advances in the MIR sensor technology, as for instance described in [10] and furthermore employed in this work [13], [14]. These novel technologies use simpler grating spectrometers, are free of any moving parts, eliminate intrinsic detector noise and are five- to ten-fold less expensive than traditional high precision FT-IR analyzers, typically employing the full spectral range ($400 - 4000 \text{ cm}^{-1}$). However, the full spectral range is rarely necessary in bioprocess monitoring as the main nutrients and products cause absorption bands in the relatively narrow spectral fingerprint region ($800 - 1800 \text{ cm}^{-1}$). Thus, apart from increasing robustness and affordability, novel technology tailors the spectrometer to this highly informative region. The technical downgrading might cause a loss in spectral quality. Yet, it is highly competitive due to increased instrumental robustness and a significant price reduction, considering that technical advances of MIR sensor technology are only at the beginning. Besides, it was demonstrated that multivariate regression is capable of dealing with the reduction of spectral selectivity and the main limitations in the modeling process remain experimental and measurement accuracy. As also demonstrated in this work, prediction accuracy based on the novel nonlinear infrared (NLIR) technology showed a prediction error of less than 15 % for several important components and hence is totally practicable for industrial large-scale fermentation monitoring.

The versatility of MIRS in combination with multivariate calibration models such as Partial -Least -Squares (PLS) has been demonstrated and reviewed in several publications as a powerful tool for monitoring fermentation processes [11], [12], [15]– [21]. However, the adoption of MIRS in biotechnology has been slow, yet with incremental progress. This might be owing to the following: IR spectroscopy combined with PLS modelling is a highly interdisciplinary field, and the building and evaluation of a strong and robust calibration model requires expertise in all the fields affecting the subject:

- 1) Process performance: The process dynamics and process interactions must be well understood and a strong, reproducible process must be in place. Besides, the signal-to-noise ratio of the IR technology in use as well as the accuracy of the reference analysis directly influence the model performance. Hence, the reference analysis must be performed by a trained person.
- 2) Chemometrics: The underlying statistics (mathematics) are complex and even though tools are available that simplify the model building process, an experienced person is necessary to evaluate the raw data and to build and to evaluate the model. Depending on the availability of software, a good level of programming skills may be useful in order to develop an in-house routine making the overall approach less time consuming and more efficient in general.

- 3) IR-spectroscopy and instrumentation: Model evaluation and optimization comes down to understanding the absorption patterns of highly overlapping spectral features and profound model understanding and evaluation is only possible when connected to the theory of IR spectroscopy as analytical method and instrumentation. This is equally important as profound process understanding.
- 4) Calibration design: The quality of the calibration set directly influences the model accuracy. When dealing with fermentation processes, the need of decoupling the correlated analyte dynamics is well known [22]. Contrarily, multi-analyte calibration design is poorly addressed in theory, methodology and software forcing researchers to construct their own custom designs or adopting known design of experiments (DoEs) which might be non-optimal for their case.

The above-mentioned points, being equally important to the subject, are rarely addressed on the same page and apart from that, no general procedures and guidelines are available on this topic of IR spectroscopy and multivariate calibration. While the potential of this combined approach is immense and demonstrated successfully ([11], [12], [15]– [21]), practice shows that the application is challenged by a lack of expertise, intensive calibration procedures and finally, often a lack of model robustness. The lack of model robustness might come down to a lack of model understanding. Multivariance is the strong advantage of IR spectroscopy and an intrinsic challenge. Especially in fermentation processes, most components are IR active and comprise the same functional groups. Hence, spectral covariance resulting from (highly) overlapping spectral features is a common reason for indirect predictions. In the frame of the subject monitoring of fermentation processes by IR spectroscopy and chemometric modeling indirect predictions remain poorly addressed. This chapter deals with the potential and challenges of IR spectroscopy and multivariate calibration. It focuses on the monitoring of glucose, glycerol, ethanol, ammonium and phosphate during a lab scale yeast fermentation process by applying a novel IR technology and PLS modeling. It addresses the challenge of indirect predictions in the framework of the employed process and points out the biological and spectral cage of covariance [23], [24]. It is important to realize that biological correlation and overlapping spectral features might be the main reason for indirect, non-robust models and that this information is not captured by the PLS model itself. In order to combine important aspects of points 1-4) the following chapter outlines important considerations when dealing with this powerful, yet highly complex strategy, evaluating model performance beyond the conventionally used parameters R^2 and Root Mean Square Error of Cross Validation (RMSECV).

2.2 Materials and Methods

2.2.1 Data collection

Fermentations of the yeast *S. cerevisiae* in the commonly used complex medium (YPD medium) were used as a case study due to its well-known (though not simple) metabolism and wide spread application in industry (e.g. baker's yeast, beer brewing, ethanol production, antibiotic production, insulin production). The data space collected for calibration and validation of the PLS models was spanned by the 6 analytes glucose, glycerol, acetic acid, ethanol, ammonium and phosphate during 4 fermentation batches. The 4 fermentations covered different process conditions with respect to the fermentation mode (batch, fed-batch) and the strain (a common CENPK reference strain and an insulin precursor (IP) producing yeast strain, described in [25]), accounting for reasonable process variation relevant to industry (see Table 2.1). Complementarily, considering the need to decouple the natural process dynamics in order to build targeted and robust PLS models for each of the analytes, batch 4 was subjected to spiking and an additional data set was prepared consisting of 15 synthetic samples containing the analytes of interest in unnatural proportions. Finally, in order to account for the presence of cells, samples from batch 4 were analyzed with and without biomass. The 6 different data sets representing a total number of 105 samples for spectral and reference analysis are listed in Table 2.1.

Table 2.1: Data set overview; CENPK refers to the yeast strain CENPK 113-7D, SIC refers to the insulin precursor producing strain as described in [25] and was kindly provided by Dina Petranovic, Chalmers University of Technology.

	Type/ No. of samples	comments
1	CENPK / 20	Batch fermentation; spectral analysis of supernatant only;
2	CENPK / 19	Fed-batch fermentation; spiked with ammonium (at time 0), phosphate and glucose (after 6 h), spectral analysis of cell-free supernatant only;
3	SIC / 19	Batch fermentation, IP producing yeast, spectral analysis of cell-free supernatant only;
4	CENPK / 16 (x2)	Batch fermentation, spectral analysis of cell-free supernatant and cell- containing samples;
5	CENPK / 16 (x2)	Spiked samples from data 4, spectral analysis of cell-free supernatant and cell-containing samples;
6	Synthetic samples / 15	Synthetic samples containing all analytes in unnatural proportions designed to minimize the correlations between the analytes

2.2.1.1 Yeast fermentation processes

Yeast fermentations were performed by cultivating the classical laboratory yeast strain CENPK-113 7D and an insulin precursor (IP) producing yeast stain (referenced as SIC and described in [25]), respectively, in YPD medium. YPD medium is a classical complex medium used for various yeast cultivations. It contained, per liter of water, 10 g yeast extract (Y, Merck France), 20 g peptone (P, Merck Mexico) and 20 g glucose (dextrose, D, Macron, USA). The YP medium was autoclaved inside the fermenter and the carbon source (glucose) was added in the form of 100 ml concentrated glucose solution (400 g/l) prior to inoculation. The process was started by adding approximately 180 ml

overnight culture via a sterile syringe. The inoculum was pre-grown in YPD medium using two 500 ml shake flasks (100 ml working volume each) at 30 °C, 180 rpm for 12-14 h. The fermentation process was run in a 2.5 liter glass bioreactor, equipped with dissolved oxygen (DO), pH and temperature probe, controlled by an Applikon ez controller. The process was started in a working volume of 2 liter and with an initial glucose concentration of around 20 g/l (110 mM). Controller and fermenter equipment were purchased from Applikon, The Netherlands. The fermentation process was run at a stirrer speed of 800 rpm, an aeration rate of 1 vvm, controlled at pH 6 using 2 M KOH and 2 M H₂SO₄, respectively, and a temperature of 30 °C. The dissolved oxygen tension (DOT) stayed above 30 % of saturation across all 4 fermentations performed, indicating that no oxygen limitation occurred during the cultivations.

Samples from the fermenter were withdrawn manually every hour by means of a sterile syringe connected to a sampling port. Each sample was split into two parts: One part was immediately filtered via a 0.2 µm filter and placed in the fridge for reference and spectral analysis after the fermentation process. The other part containing biomass, was simultaneously analyzed for the biomass content by traditional optical density measurement at 600 nm (OD₆₀₀, as described in 1.2.3 Off-line glucose measurements in yeast fermentation samples) and by spectral analysis in case of data set 4 and 5. In case of data set 5, spiking was performed on the cell-containing sample which was then split and analyzed as described.

In the case of data set 2, an initial ammonium level of 100 mM was adjusted by adding a proper volume of a 2 M ammonium stock solution, pH 6 (prepared from ammonium chloride, Sigma, USA). After 6 h, 100 ml of a 2 M phosphate stock solution, pH 6 (prepared from H₂PO₄, in 4 M NaOH (Sigma, USA)) and 100 ml of glucose stock solution (400 g/l) was added. Naturally, YPD medium does only contain low levels of inorganic phosphate and ammonium of around 15 mM each. The addition of up to 100 mM ammonium and phosphate was done to meet the calibration range previously used for PLS modeling of phosphate and ammonium (see appendix 1). Glucose was added in order to prolong the exponential growth phase and thus vary the process dynamics in a fed-batch like fashion.

2.2.1.2 Design of synthetic samples

The fermentation parameters analyzed are inherently trapped in a biological cage of covariance, meaning that they are highly correlated with each other due to the microbial metabolic network activity from which they result. Thus, a very first CENPK batch fermentation (not included in the data sets for spectral analysis) was performed in order to have a reference with respect to the concentration profiles expected for the 6 analytes and the correlations between each other. Only the batch phase was considered for this reference data set determined by the depletion of glucose. The batch phase generally covers the concentration variation of all analytes completely. [Figure 2.2](#) shows the concentration profile of the 6 analytes under study ([A](#)), their respective correlation factor (r value, [B](#)) and the correlation profile between each other([C](#)).

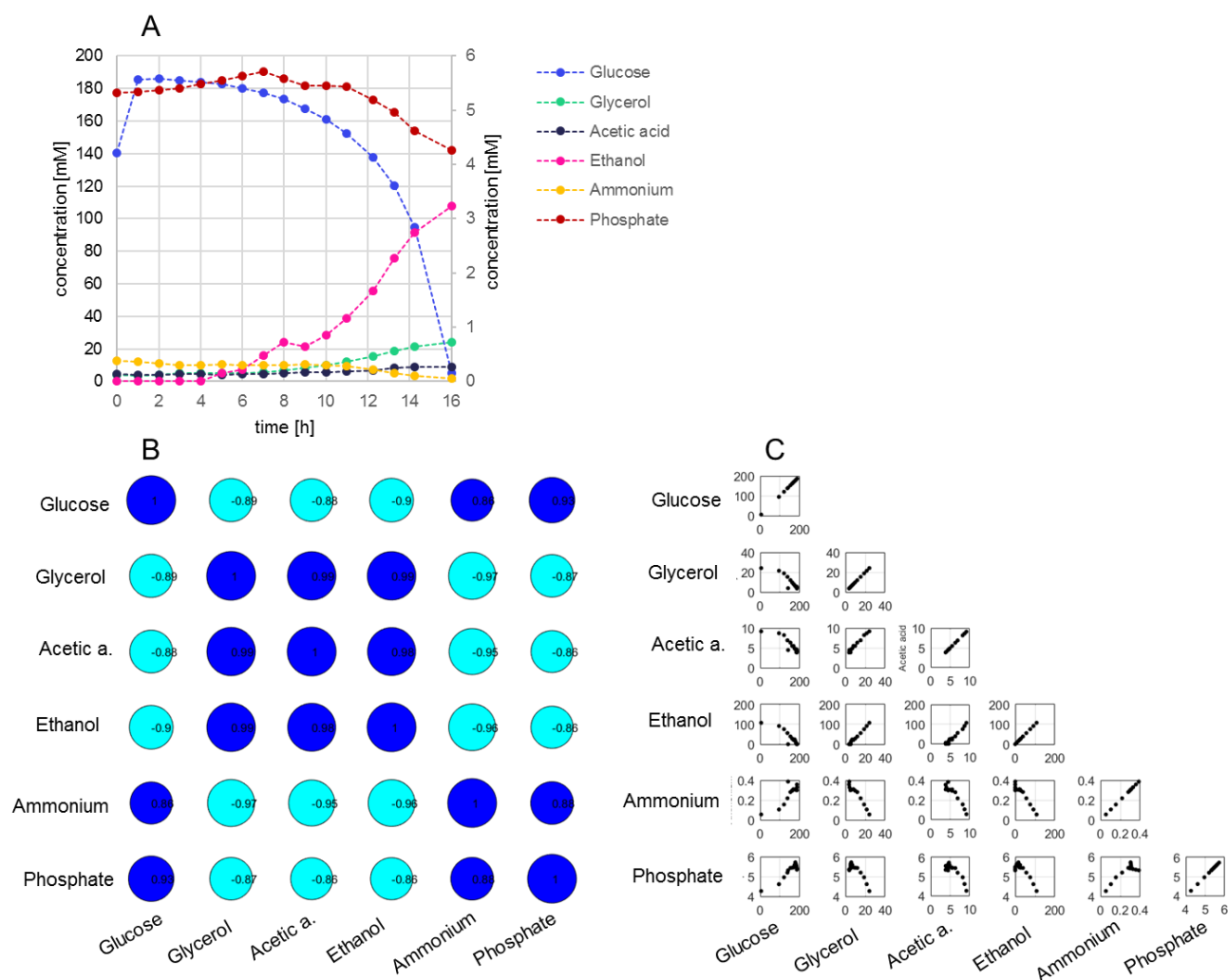


Figure 2.2: A) Analyte profiles over the course of the CENPK reference batch fermentation. The end of the batch phase is characterized by glucose depletion after 16 h. The left y- axis refers to the concentrations of glucose, glycerol, acetic acid and ethanol. The right y- axis refers to the concentrations of ammonium and phosphate. B) Bubble plot of the correlation coefficient (r) values between the components of interest ranging from -1 to 1. The dark blue color indicates a positive correlation while the light blue color indicates a negative correlation between two components. The bubble size indicates a strong (big bubble, r close to 1 or -1) or weak (small bubble, r close to 0) correlation between two components. The diagonal can be seen as a reference for the bubble size of a r value equal to 1 (100 % correlated). The average correlation between the components is 0.92 and with that very high. C) Each subplot reveals the correlation profile of two components (concentration in [mM] against concentration in [mM]). The respective r values of each correlation profile are visualized in B).

Figure 2.2 A shows a typical fermentation profile of a yeast batch fermentation. Glucose is transformed into ethanol as well as glycerol and acetic acid as (minor) by-products. Detailed information about the yeast metabolism can be found elsewhere [26], [27]. It can also be observed that the phosphate and ammonium level in the complex medium (as part of the yeast extract and peptone) is very low and barely varies over the course of the fermentation. This can be explained by the preferred use of organic phosphorous (P) and nitrogen (N) sources over the inorganic ions such as phosphate and ammonium. Many organic building blocks are brought into the medium by yeast extract and peptone and it is assumed that they are preferably used by the microorganism as P and N sources compared to the inorganic ions. The correlation coefficient between the components is high, on average an r value of 0.92 was obtained for the different analytes making it impossible to model the components independently from each other by PLS. In order to overcome this issue, 15

synthetic samples were designed based on the reference analyte concentration profile shown in Figure 2.2 A. The concentration values of each analyte over time can be seen as the design matrix constituted by 6 columns (6 analytes) and 15 rows (15 time points during the fermentation). In order to create samples with uncorrelated analytes but realistic process values, the concentration values in each column were randomly permuted by a built-in Matlab script kindly provided by Frans v. d. Berg, University of Copenhagen, Department of Food Science. Thus, 15 samples were created showing a combination of concentration values, each a true process value as such, but in combination completely unrealistic and uncorrelated. In order to increase the leverage of the data set, the maximal concentration value found for each component was as much increased as considered to be relevant. Finally, all concentration values were rounded to the nearest 5 mM for practical reasons. The composition of the synthetic samples as well as the correlation profiles between the different analytes of interest, now decoupled in this data set, are shown in Figure 2.3 A-C.

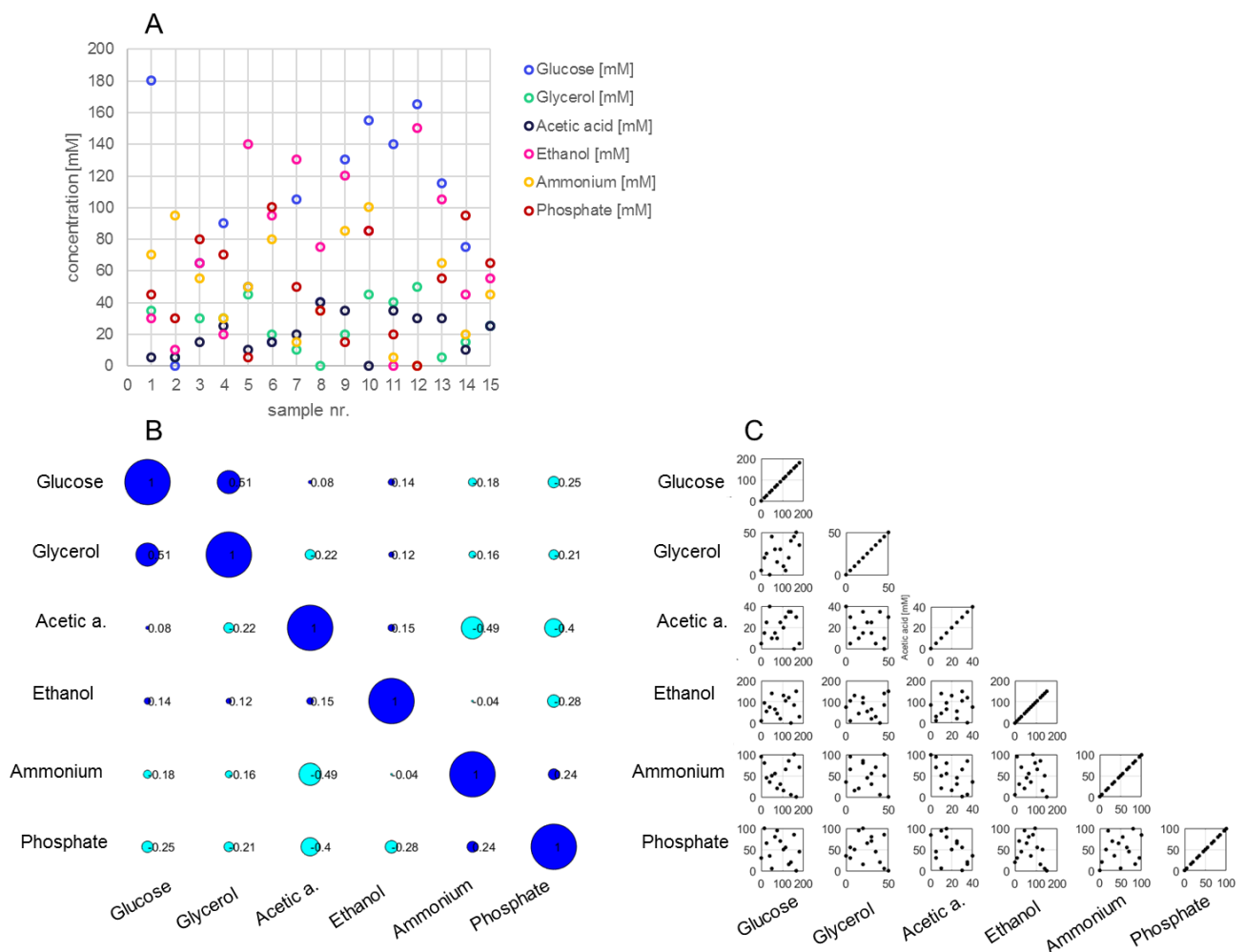


Figure 2.3: A) Permuted sample composition of the 15 synthetic samples. B) Bubble plot of the r values between the components of interest. The dark blue color indicates a positive correlation, while the light blue color indicates a negative correlation between two components. The bubble size indicates the strength of the correlation and can be found as significantly decreased compared to Figure 2.2 B. The diagonal can be seen as a reference for the bubble size of an r value equal to 1 (100 % correlated). The average correlation between the components is 0.23 and with that the different analytes can be considered as uncorrelated. C) Each subplot reveals the correlation profile of two components (concentration in [mM] against concentration in [mM]). The respective r values of each correlation profile are visualized in B). The subplots clearly show that a coherent trend as found in Figure 2.2 C is no longer present.

In the synthetic samples, the correlation between the different analytes was significantly reduced to an average r value of 0.23 (experimental outcome 0.33). In this data set, the analytes can be considered as uncorrelated. Including the synthetic samples inside the model calibration set ensured that the process correlations were decoupled and the different components were more likely to be modeled independently from each other (from a process point of view).

2.2.1.4 Preparation of synthetic samples

The synthetic samples (data set 6 in [Table 2.1](#)) were prepared based on matrix spiking creating a total sample volume of 15 ml each, adjusted to the operational pH of 6. The matrix (YP medium at pH 6) was spiked with a proper volume of a 2 M stock solution of each component according to the concentration scheme shown in [Figure 2.3 A](#). The composition of each sample is summarized in the supplementary material, Table S2.1 and the corresponding spiking scheme is presented in Table S2.2. Glucose, acetate and ammonium stock solutions were prepared by dissolving the appropriate amounts of solids (glucose (dextrose), Macron, USA, sodium acetate trihydrate, Sigma USA, ammonium chloride, Sigma USA) in purified water. Stock solutions of phosphate, glycerol and ethanol were prepared by adding the appropriate amount of 85 % phosphoric acid (Aldrich, China), glycerol (VWR chemicals, Belgium), and ethanol (Sigma Aldrich, Germany), respectively, to purified water. All solutions were adjusted to pH 6 by adding a few drops of 4 M HCl or 4 M NaOH. It was chosen not to use the operational pH reagents KOH and sulfuric acid, as the sulfate ion is IR active, whereas the ions H^+ , Na^+ and Cl^- are not. The O-H group is IR active, however, its absorption bands, and respectively, generally the absorption caused by water, were excluded from the spectral range selected for PLS modeling. As the stock solutions were highly concentrated, a minor amount of spiking solutions was added compared to the matrix volume and the matrix was not diluted more than 1.25 times, which was considered to be negligible. It must be noted, that, also during a standard fermentation process, the initial medium, or matrix, respectively, gets diluted in a comparable manner due to the addition of acid and base.

In order to investigate the spectral fingerprint of the pure components, solutions in Milli Q water containing only one component were prepared. Each component was prepared in a concentration of 10, 50, 100 and 200 mM in order to be able to follow the concentration gradient in the spectra.

2.2.1.5 Spiked samples

On the one hand, spiked samples can be used to disturb the process dynamics and thus decouple the correlations between the components. On the other hand, due to their varied content, they present an excellent measure for model robustness when used as external validation. Given that their concentration values are inside the calibration space, spiked samples would not be predicted accurately if the model was built upon indirect correlations. Or in other words, their prediction would only be accurate when the model was built upon direct correlation based on distinct spectral features being directly connected to the analyte of interest.

The spiking pattern for the spiked samples (data set 5 in [Table 2.1](#)) was developed based on a reference fermentation, as the synthetic samples. For the spiked samples the respective reference batch was data set 1 ([Table 2.1](#)). The first 15 samples (0-14 h) were chosen as design matrix constructed by the 6 analytes (columns) and 15 time points (rows). As for the synthetic samples, an

uncorrelated spiking scheme was developed by random permutation of the row values. As spiking was applied to true fermentation samples as background, meaning that no 'clean' background was present as was the case for the synthetic samples, the reference data set was subtracted from the spiking scheme obtained by permutation. If the reference concentration value (data set 1) was higher than the concentration value aimed for in the spiked sample, the value in the spiking scheme was set to zero and nothing was added for the respective component. Otherwise, the difference between reference data point and designed data point was added to the sample, in order to obtain the actual design point. This procedure should keep the spiked samples approximately inside the concentration space spanned by the reference data set 1. Ammonium and phosphate were added up to 100 mM in order to incorporate a greater variability relevant to other processes. As for the synthetic samples, the final concentration value was rounded to the nearest 5 mM. The correlation coefficient in the designed data set was 0.246 resulting in an experimental outcome of 0.127. Spiking was performed right after sampling. Each sample taken represented a time specific background matrix to which the analytes were added (spiked) according to the spiking scheme. The final volume of each spiked sample was 10 ml. Spiking solutions were prepared as described in the previous section, 2.2.1.4 Preparation of synthetic samples. The spiking scheme is presented in the supplementary material, Table S2.3.

2.2.3 Reference analysis

All measurements were performed in duplicates.

2.2.3.1 Glucose and metabolite quantification

Glucose, glycerol, acetic acid and ethanol were quantified by High-Performance-Liquid Chromatography (HPLC). HPLC analysis was performed on an Ultimate 3000 Dionex HPLC system (Sunnyvale, USA) using an Animex HPX 87 H column, 300 x 7.8 mm (BIORAD, Denmark) operated at 50 °C equipped with Refract Max 520 refractive index (RI) detector. The column was operated with 5 mM H₂SO₄ in purified water as mobile phase and at a constant flow rate of 0.6 ml/min, injecting 5 µl of sample for analysis. Samples were filtered and acidified (950 µl sample + 50 µl 5M H₂SO₄) prior to analysis. The acidification of samples was necessary due to the ion exchange principle of the column used. Sample filtration is a general requirement when analyzing cell (particle) containing samples by HPLC to avoid the blocking of the small capillaries. Sample quantification was done with the software Chromeleon 6.8.

2.2.2 Ammonium quantification

Ammonium concentrations were determined after diluting to a proper concentration using the Ammonium Cuvette -Test 47-130 mg/L NH₄-N (LCK 302) from Hach Lange. In this assay, ammonium ions react at pH 12.6 with hypochlorite and salicylate ions in the presence of sodium nitroprusside as a catalyst to form indophenol blue (measured at 694 nm). Samples were analyzed in the Hach Lange Spectrophotometer DR 3900.

2.2.3 Phosphate quantification

Phosphate concentrations were determined after diluting to a proper concentration using the Phosphate Colorimetric Assay KIT from Sigma Aldrich (MAK030). During this assay inorganic phosphates (Pi) react with a chromogenic complex, which results in a colorimetric (650 nm) product proportional to the amount of phosphates present. Samples were analyzed on a 96 well plate (Costar Assay Plates, REF. 3370) with the spectrophotometer Multiskan Go from Thermo Scientific.

2.2.3 Spectroscopic analysis

IR spectra were acquired in duplicates on two different instruments. One instrument can be considered as a commercial standard lab-instrument (ABB Bomem MB100) based on the ATR-FT-IR principle. The other instrument was a novel technology named NLIR, described in section 2.2.3.2.3 NLIR technology. The NLIR instrumentation used was a prototype in development kindly provided by NLIR Aps (Nonlinear Infrared Sensors, Farum, Denmark). PLS models were developed based on the spectral data acquired with both instruments and compared with respect to their performance, represented by the relative Root Mean Square Error of Prediction (RMSEP in %).

2.2.3.1 ATR-FT-IR (standard instrument)

Infrared spectra were acquired off-line on a FT-spectrometer (ABB Bomem MB100) equipped with a Globar silicon Carbide (SiC) radiation source collimated to a 2.5 cm diameter beam and a deuterated triglycine sulfate (DTGS) detector. The measurements were performed in attenuated total reflectance (ATR) mode, by putting 800 μL of each sample on the ATR cell (ZnSe crystal, 45°C) and recording a full spectrum between 750 and 4000 cm^{-1} with 64 scans at a resolution of 8 cm^{-1} . The background was obtained on an empty ATR groove against air using 128 scans. In between the samples, the ATR groove was cleaned with water and 70 % ethanol and dried out with spectrophotometric tissues. Subsequently, only the region between 900 and 1550 cm^{-1} was kept for further analysis. This region refers to the region facilitated by the NLIR instrument. Besides, it is within the so called fingerprint region, which is rich on spectral information linked to the different functional groups present in biological samples.

FT-IR based instruments inherently produce data linear in frequency, and thus, the spectra were measured in wavenumbers [cm^{-1}].

2.2.3.2 NLIR- Nonlinear IR (prototype)

2.2.3.2.1 Off-line measurements

Infrared spectra were acquired with the NLIR instrument, equipped with a 1200 K silicon nitride filament as infrared light source and a grating spectrometer with a silicon-based CCD line array as detector. The measurements were performed in attenuated total reflectance (ATR) mode, by putting 800 μL of sample on the ATR cell (ZnSe crystal, 45°C, 12 bounces) and recording a spectrum between 800 and 1800 cm^{-1} with 100 scans at a resolution of 10 cm^{-1} and an exposure time of 40 ms. The background was obtained on an empty ATR groove against air using 100 scans. In between the samples, the ATR groove was cleaned with water and 70 % ethanol and dried out with spectrophotometric tissues. Subsequently, only the non-noisy region, between 900 and 1550 cm^{-1} , was kept for further analysis.

Compared to the FT-IR instrument in section 2.3.1, the NLIR instrument inherently produces data linear in wavelength [nm]. Wavelength was transformed into wavenumbers [cm^{-1}] to stick to the convention that Mid-IR spectra are shown as a function of wavenumbers [cm^{-1}].

The correlation between wavenumber (ν) and wavelength (λ) is given in Equation 2.1

$$\nu [\text{cm}^{-1}] = \frac{10^7}{\lambda [\text{nm}]} \quad \text{Equation (2.1)}$$

2.2.3.2.2 On-line measurements

As the ultimate goal was to develop an online set-up for bioprocess monitoring by IR spectroscopy, mid-IR measurements with the NLIR instrument were conducted on-line on a final yeast fed-batch fermentation. Respectively, the fermentation was spiked with 50 ml of a 400 g/L glucose solution, after 9 h of the fermentation course. For this set-up, the stationary ATR (ZnSe crystal, 45°C, 12 bounces) was exchanged with an ATR crystal (ZnSe crystal, 45°C, 10 bounces) designed as a flow through cell (HATR Flow-Through Cell, PIKE Technologies, USA, [9]). Inlet and outlet of the ATR flow-through cell were luer connectors, thus it could be easily connected to the fermenter. The fermentation fluid was recirculated from a sampling port equipped with a 20 μm stainless steel filter cap (sample screen 20 μm , Applikon, The Netherlands) to prevent the sampling of air, to the ATR flow-through cell, and via a pump (Ismatec Reglo ICC, Ismatec, Germany) recirculated back to a separate fermenter inlet. The volume of the flow through cell was 500 μl , considering inlet and outlet volumes of the cell the total volume was approximately 1 ml. The recirculation loop was operated with a flow rate of 5 ml/min. Taking the dead volume inside the tubing to the flow-through-cell into account, an exchange of fermentation broth inside the flow-through cell was ensured every minute. IR spectra were automatically recorded every 10 minutes while a manual sample for reference analysis was taken every hour via a second sampling port. Thus, every 6th MIR measurement was complemented with off-line reference data.

2.2.3.2.3 NLIR technology

As the NLIR instrument is based on novel technology (featured in detail in [13], [14], [28]), some main technical considerations regarding the core technology are given in the following paragraph. Unlike in classical FT-IR spectroscopy used in the standard instrument in 2.3.1, where the most important part is represented by the interferometer (described in [8]), the core technology in the NLIR instrument is the laser cavity illustrated in Figure 2.4. The IR light passes through the ATR and enters the laser cavity in which a process called upconversion takes place. With respect to Figure 2.4, the light processing in the laser cavity can be described as follows:

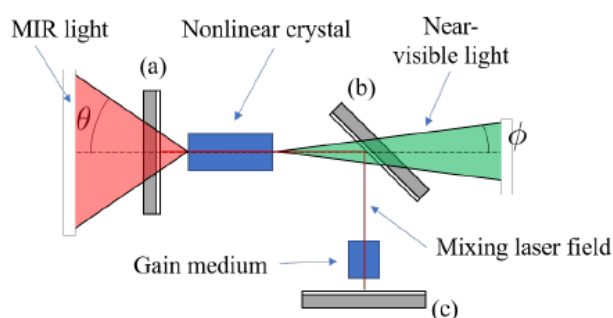


Figure 2.4: Laser cavity: MIR light (6.0 – 11.5 μm) that has interacted with the sample on the ATR enters from the left and exits to the right after upconversion to near-visible light (904 – 963 nm). Mirror (a) transmits MIR wavelength, (b) transmits visible wavelength and all three mirrors (a), (b), (c) reflect the mixing laser wavelength (1064 nm). The figure was borrowed with permission from NLIR Aps.

MIR light enters the laser cavity through the dichroic mirror (a) transmitting the IR wavelength of interest (approximately 6.0 – 11.5 μm corresponding to 870 – 1666 cm^{-1}) while reflecting the others (noise). The incoming infrared light is combined with a 1064 nm high power mixing laser of

approximately 3 W and focused into the non-linear crystal (AgGaS₂). Inside the non-linear crystal, the upconversion from MIR to near visible light (904 – 963 nm) takes place by sum frequency generation (SFG). Inside the non-linear crystal, a MIR photon and a mixing laser photon (1064 nm) annihilate and create a near-visible wavelength photon based on energy conservation (Equation (2.1) and (2.2)). After the non-linear crystal, the 1064 nm laser, conserved in the laser cavity by reflection, and the generated near visible light are separated by another dichroic mirror (b) through which only the visible light can leave the laser cavity. The laser light is kept inside being trapped (reflected back and forth) between the mirrors (a), (b), and (c).

The generated near visible light is sent into a grating spectrometer with a silicon-based CCD line array as detector where all wavelengths in the near-visible spectrum (904 – 963 nm) are recorded simultaneously.

Generally, light of the MIR region is difficult to detect due to Planck radiation of detectors with finite temperature. Compared to the conventional technology described as standard, where IR light is directly detected by an IR detector, this technology converts incoming IR light that has interacted with the sample to near-visible light. Thus, a standard silicon based visible light detector (CCD / CMOS) can be used for detection. These detectors are efficient, free from intrinsic thermal background noise (compared to a conventional IR detector that emits IR light itself due to its own temperature (black body radiation), thus, increasing the noise) and cheap. The spectral information in the upconverted light is completely conserved because of energy conservation. The equation for photon energy and energy conservation is described in equation (2.2) and (2.3),

$$E = hf = \frac{hc}{\lambda} \quad \text{Equation (2.2)}$$

$$\frac{1}{\lambda_{up}} = \frac{1}{\lambda_{MIR}} + \frac{1}{\lambda_{1064}} \quad \text{Equation (2.3)}$$

where E is energy, f is the frequency, h is Planck's constant and c is the speed of light. λ_{up} , λ_{MIR} and λ_{1064} are the wavelength of upconverted light, incoming MIR light that has interacted with the sample and the mixing laser, respectively, used for upconversion.

Visible light detectors are orders of magnitudes better in terms of detectivity than any MIR light detector, which clearly highlights the advantage of the upconversion process. The combination of frequency (energy) upconversion and visible light detector is the key to fast and efficient MIR wavelength detection in the NLIR spectrometer. Furthermore, the optical core of this technology does not contain any moving parts (a moving mirror is typically part of an interferometer) resulting in increased robustness against mechanical vibrations. The term nonlinear (NL) IR results from the field of nonlinear optics to which the theoretical background of this technology belongs.

It must be highlighted that spectroscopic studies with this instrument have rarely been conducted on fermentation samples and, to my knowledge, this case is the first to be published within the bioprocess field.

2.2.4 Chemometric modelling

The PLS algorithm describes the relationship (calibration model) between the analyte concentrations (**y**) and the IR spectrum (data matrix **X**). It is suitable for handling the regression of **y** on numerous, highly correlated variables as presented by IR spectra in general. PLS approximates the spectral data by a smaller number of so-called latent variables. The latent variables are linear combinations of the original spectral values describing the variation in **X** covarying with the information in **y**. They are obtained by maximizing the covariance between **X** and **y** using a least square approach. The

calibration yields a PLS model which can be used for the prediction of analyte concentration based on spectral data only. PLS models are evaluated by the correlation coefficient between reference and predicted values (R^2) and the Root Mean Square Error of Prediction (RMSEP). The RMSEP is calculated according to Equation (2.4). Within this work, the unit of the RMSEP is mM.

$$\text{RMSEP [mM]} = \sqrt{\frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{N}} \quad \text{Equation (2.4)}$$

with N equals the total number of samples in the validation set, y_i equals the reference value of sample i and \hat{y}_i equals the predicted value of sample i . RMSEP values are based on external validation, that is to say data in the validation set was not included in the calibration set. This yields a more realistic idea of model performance as the traditionally used RMSECV (root mean square error of cross validation). In many works, the RMSECV is considered for estimating the model performance on future samples. Cross validation is a broadly used internal validation procedure used as an estimate of model performance on new samples. This method describes a loop, where predefined data set segments are left out one after the other from the modelling phase. A model is established on the remaining data segments and this model is then applied to the excluded data segment. In each cross validation loop the predictions of the excluded data point are collected. These predictions are used against the reference values to estimate the RMSE. For the final RMSECV of the model the prediction of all segments is reported. This often gives an overoptimistic idea of model performance as it does not account for batch-to-batch variability and indirect predictions.

The relative RMSEP in % gives a more profound idea of model performance than the total RMSEP calculated in (4). It relates the error to the actual concentration range predicted, a range that may be very different for the analytes of interest. It is obtained by Equation (2.5)

$$\text{RMSEP [\%]} = \frac{\text{RMSEP [mM]}}{\Delta y_{\text{val}} [\text{mM}]} * 100 \quad \text{Equation (2.5)}$$

with Δy_{val} equals the concentration range of the respective component used in the validation set.

2.2.4.1 PLS calibration and PCA analysis of spectral data

Chemometric analysis was performed in MatLab R2017b (MathWorks, Natick, MA), using both the PLS_Toolbox (Eigenvector Research Inc., Manson, WA) and in-house written routines.

PLS calibration models [29], [30] were built using the SIMPLS algorithm and a 10-fold cross-validation procedure that kept all spectral replicates in the same subset. The spectral raw-data of each calibration set was subject to mean-centering and the first derivative using the Savitzky-Golay algorithm [31] (window-size of 25 data points and a 3rd polynomial order, first derivative (25, 3, 1)). Savitzky-Golay applying the first derivative was chosen after comparison with other common spectral preprocessing methods, namely Standard Normal Variate (SNV) [32], second derivative and smoothing [33]. Spectral derivation and smoothing is generally obtained using the Savitzky-Golay algorithm. The PLS models obtained after preprocessing with the first derivative yielded the overall best results. Hence, Savitzky-Golay applying the first derivative was chosen as spectral pre-processing, after mean-centering.

The Principle Component Analysis (PCA) models [34], [35] were built on the whole data set of samples. In a first case, the spectral data was mean-centered. In a second case, the data was mean centered and subjected to the first derivative, aiming at a PCA model describing the spectral variance on the preprocessed data as used for PLS calibration.

2.3 Results and Discussions

2.3.1 Fermentation data

The fermentation profiles of data sets 1-4 as well as the spiked and synthetic samples (data set 5 and 6) as described in Table 2.1 are shown in Figure 2.5. Glucose, glycerol, acetic acid and ethanol were quantified by HPLC with an overall standard deviation < 2 %. Phosphate and ammonium were quantified by colorimetric assay kits with an overall standard deviation < 5 %. If the given confidence interval was exceeded in case of some data points, it was marked in the respective plots concerning PLS model calibration and validation (see Figures in section 2.3.4.3 Detailed presentation of Model 3a). The analyte profiles are split according to the 6 components of interest in order to visualize differences in the concentration profiles within each component. This was important for the subsequent evaluation of data sets chosen for PLS calibration and validation, respectively.

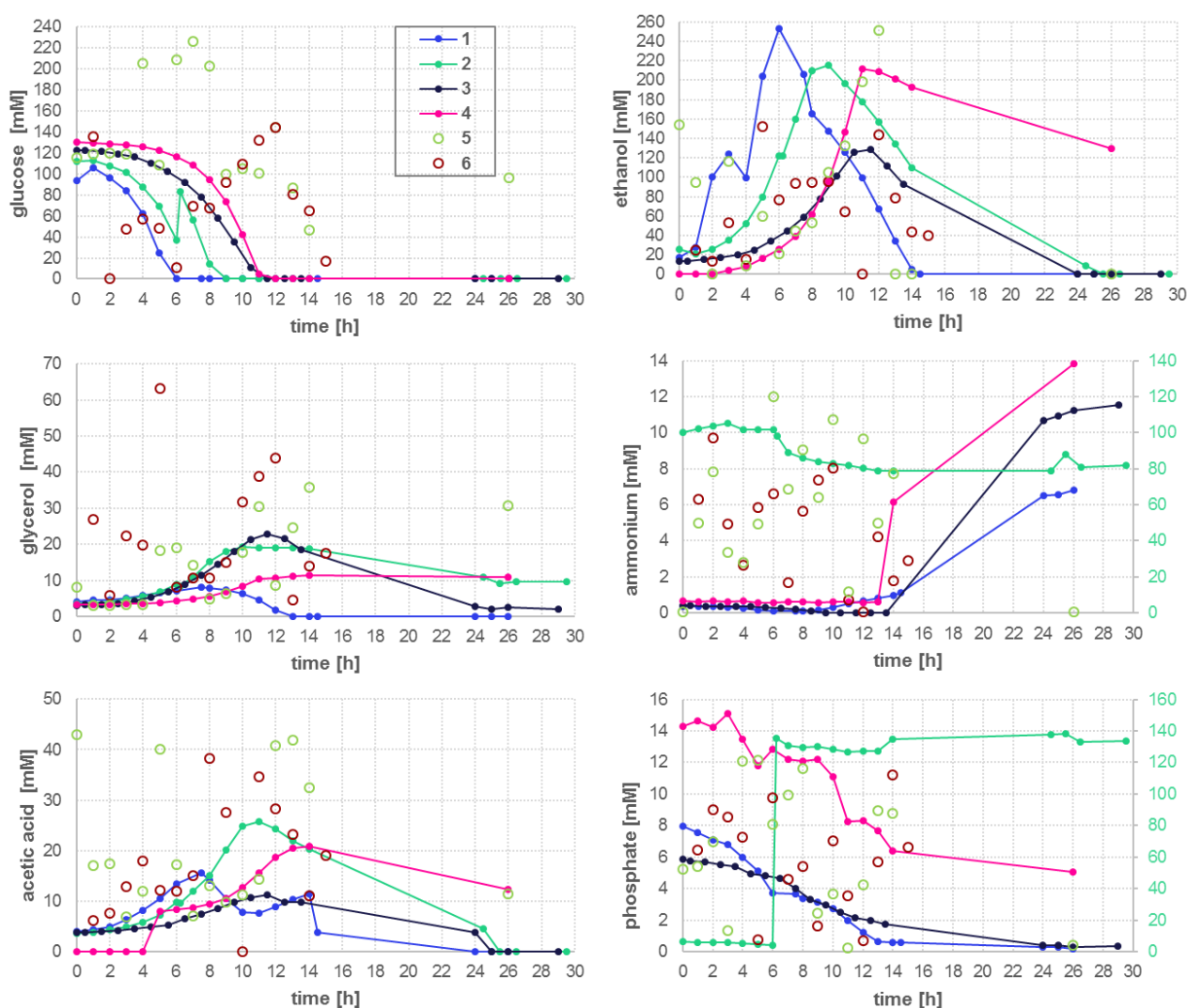


Figure 2.5: Fermentation profiles of the data set 1 - 4 as well as spiked and synthetic samples are described in Table 2.1. Glucose, glycerol, acetic acid and ethanol were quantified by HPLC with an overall standard deviation < 3 %. Ammonium and phosphate were quantified by colorimetric assay kits with an overall standard deviation < 5 %. The legend shown for the analyte glucose holds for all the other components and the numbering in the legend refers to the data set number as outlined in Table 2.1. In case of the data for ammonium and phosphate, the right y-axis refers to sample set 2, 5 and 6 (spiked and synthetic samples), and the left y-axis refers to data set 1, 3 and 4 (unmodified batches).

Figure 2.5 was evaluated from both, a metabolic and a modelling perspective. From a metabolic point of view, the data shown in Figure 2.5 assemble a typical yeast fermentation process. Glucose is consumed and mainly transformed into ethanol and biomass (Biomass data not shown). Next to ethanol, acetic acid and glycerol are produced and secreted into the medium due to overflow metabolism. When the glucose is depleted, the so-called diauxic shift happens, and the cells start to grow on ethanol as well as on acetic acid and glycerol as second carbon- and energy source. Hence, the fermentation presents a common two-phase process, starting with phase 1, in which glucose is consumed and metabolites are produced, followed by phase 2, the consumption of metabolites that were previously produced. Detailed information about the yeast metabolism can be found in [26] and [27].

From a modelling point of view, three findings are striking with respect to Figure 2.5. First of all, the variability between the different batches was rather high, making the choice of calibration and validation set even more crucial, as generally speaking, accurate prediction can only be expected inside the calibration space. Data points outside the calibration space are expected to result in poor predictions as the PLS model is then forced to extrapolate. In case of complex process dynamics, resulting in complex data matrices as the IR spectra, extrapolation is likely to be error prone. Secondly, the variability space (concentration range) was much higher in the case of glucose and ethanol (disregarding ammonium and phosphate in the spiked batches and synthetic samples, data set 2, 5 and 6), than in case of the analytes acetic acid and glycerol. Not considering the spiked and synthetic samples (data set 5 and 6) glucose and ethanol varied in a range of 0 – 144 mM and 0 – 253 mM, respectively, whereas glycerol and acetic acid varied only up to 36 mM and 43 mM, respectively. Thirdly, the level of ammonium and phosphate in the complex medium was found to be very low, only around maximal 1 mM and 15 mM, respectively. It could furthermore be observed that the initial ammonium level measured in the fermentation medium was consistent across the data sets (disregarding the spiked batch 2). Contrarily, the initial amount of phosphate varied broadly across the batches performed, and measured concentrations are between 1 mM and 14 mM (disregarding the spiked and synthetic samples, data set 5 and 6). The differences in the initial phosphate levels in the fermentation broth might be a result of variations in the initial amounts of yeast extract and peptone, weighed in for media preparation. Other reasons might be variations in the times the fermenter stayed inside the autoclave following autoclaving (over night or only approximately 2-3 h) and variations in the yeast extract used (batch-to batch variability of complex raw material). The concentration trend of both components is considered in the following. Inorganic ammonium seems to be untouched by the microorganism during the batch phase and increased after glucose depletion, eventually due to cell lysis. This advocated that nitrogen needed for cell growth was taken up in the form of organic nitrogen species. These are plenty available in a complex medium in form of small peptides and amino acids. Inorganic phosphate seemed to undergo an actual consumption by the microorganism, especially during the batch phase. However, compared to the organic analytes, glucose, ethanol, glycerol and acetic acid, the inorganic species ammonium and phosphate were only present in very low concentrations and underwent comparably only a minor variability.

2.3.2 Spectral analysis

2.3.2.1 Pure component spectra

IR radiation is absorbed by exciting fundamental vibrations of molecular bonds expressing a change in dipole moment. Thus, highly polar bonds result in absorption bands of higher intensity compared to less polar bonds. The best example for a highly polar bond resulting in an immense spectral fingerprint is the O-H bond of a water molecule. Water absorption bands dominate the IR spectrum of aqueous solutions and appear at ca. 1630 cm^{-1} due to O-H bending and between $3470 - 3200\text{ cm}^{-1}$ due to H-O-H stretching vibrations. Generally, organic components are to a large extent composed of the atoms C, N, O, P and H. Due to the difference in electronegativity of these atoms, bonds between them are inherently polar and prone to IR absorption. This also holds for the inorganic ions ammonium and phosphate composed of N-H and P-O(H) bonds, respectively. Figure 2.6 shows the molecular structure of all 6 analytes and functional groups expected to contribute to the IR spectrum due to their polar nature are highlighted.

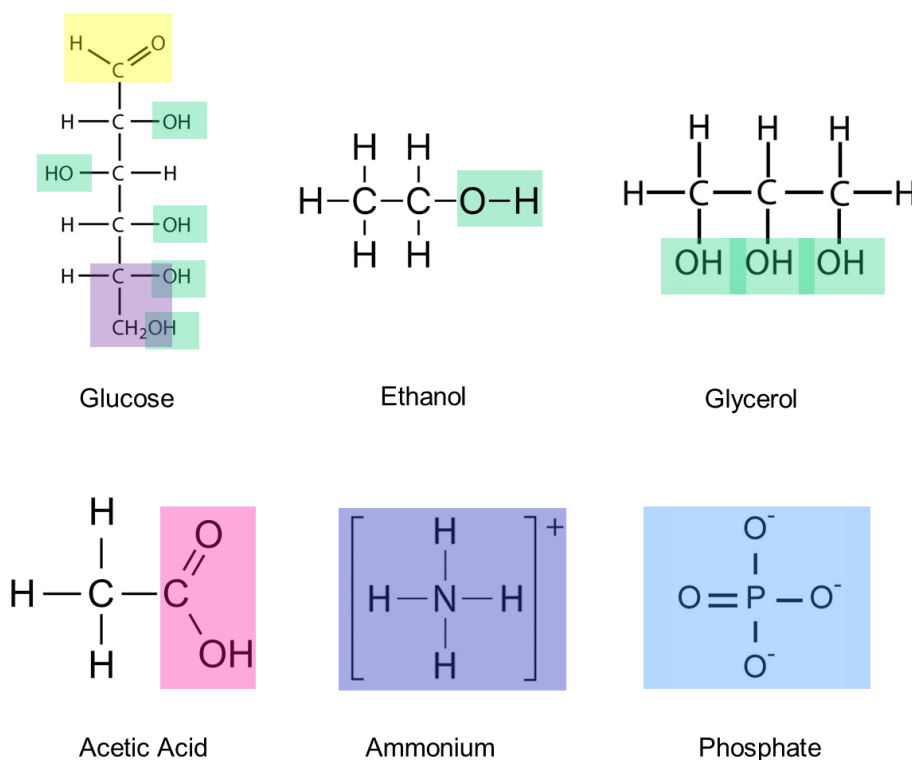


Figure 2.6: Molecular structure of the 6 analytes glucose, glycerol, acetic acid, ethanol, ammonium and phosphate. Functional groups contributing to IR absorption are highlighted in the organic components. Ammonium and phosphate must be seen as an individual functional group found also in biomolecules e.g. as amine (-NH_2) or phosphate group.

Mainly, C-O(H) stretching vibrations of alcoholic groups (green in Figure 2.6) are expected to be important to the three analytes glucose, glycerol and acetic acid. The carbonyl group of glucose (yellow in Figure 2.6) is expected to excite a different vibrational mode. Furthermore, the carboxyl group of acetic acid, as well as ammonium and phosphate as such can be seen as individual moieties causing a different spectral variation each. They are expected to cause absorption bands due to the CO_2^- symmetrical stretching mode of the carboxyl group in acetic acid (pink in Figure 2.6) and N-H as well as P-O stretching vibrations in ammonium and phosphate, respectively. Already at this point,

simply by studying the molecular composition of the 6 analytes, it becomes evident that (at least) the three components glucose, glycerol and ethanol will show overlapping spectral features due to their common alcoholic group describing one of the main structural moieties of these molecules. Stretching vibrations involving the relevant atoms C, O, N, P, H characteristically occupy the so called fingerprint region of mid-IR spectra which is typically ranging from 800 to 1800 cm^{-1} .

In order to complement the theoretical considerations, the pure component spectra of all 6 analytes were investigated and their spectral fingerprint was assigned (Figure 2.7). The spectral region examined was reduced to 900 – 1550 cm^{-1} in order to exclude the strong water absorption band at 1630 cm^{-1} .

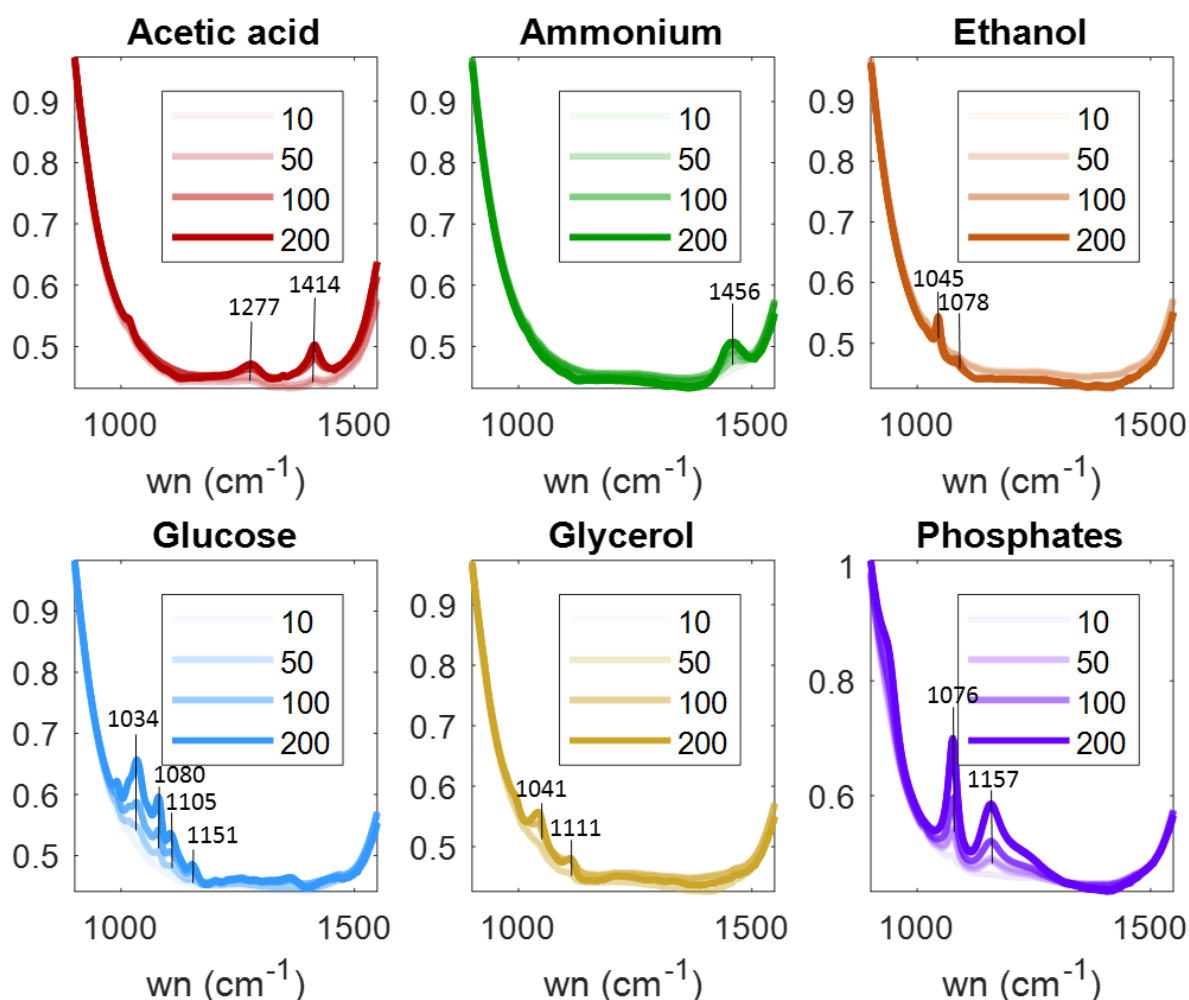


Figure 2.7: IR spectra of pure components in Milli-Q water, acquired with the standard FT-IR instrument. Each spectrum shows the absorbance as a function of wave number (wn) in cm^{-1} of 4 different analyte concentrations as indicated in the legends of each figure (10, 50, 100 and 200 mM).

Based on Figure 2.7, the major peaks for each component were assigned with wavenumber and connected to the respective vibrational mode that was excited. The results were compared to literature. The findings are summarized in Table 2.2.

Table 2.2: Peaks identified for each component in the fingerprint region (900 – 1550 cm⁻¹), based on Figure 2.7. The absorption bands found were compared with literature and the vibrational mode excited was assigned.

component	spectral peaks identified [cm ⁻¹]	excited vibration, reference
glucose	1034 1080 1105 1151	C-O stretch C-O stretch C-O stretch C-O-C stretch [36]
glycerol	1041 1111	C-O stretch C-O stretch [7]
ethanol	1045 1078	C-O stretch C-O stretch [36],[7],[10]
acetic acid	1277 1414	C-O ₂ symmetrical stretching (carboxyl gr.) C-O ₂ symmetrical stretching (carboxyl gr.) [36], [37]
ammonium	1456	Triply degenerated bending mode of the tetrahedral ammonium ion [38], [39], [40], [41]
phosphate	1076 1157	P-OH stretch P-O2 stretch [42], [43]

Additionally, to strike the point of overlapping spectral features visually, the pure component spectra of the 200 mM solutions are shown in an overlaid fashion in [Figure 2.8 A](#). The specific molecular absorbance of each component was calculated based on the highest peak identified in the 200 mM solution and is shown in [Figure 2.8 B](#). A more representative way to calculate the molecular absorbance would be to consider the whole area under the respective peaks. However, considering the absorbance of the highest peak identified gives a realistic estimate of the IR-activity (molecular absorbance) of each component.

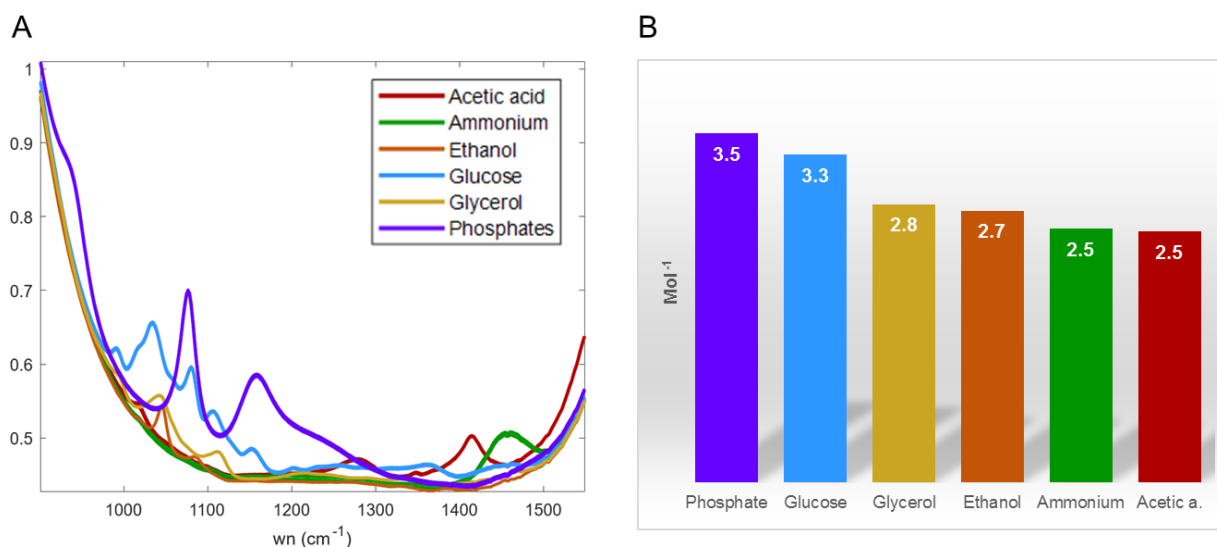


Figure 2.8: A) Pure component spectra of the 200 mM solution in an overlaid fashion. B) Molar absorption of the highest peak identified for each component in Mol^{-1} .

The broadly overlapping spectral features become obvious in Figure 2.8 A. The left-hand side of the fingerprint region (1000 -1200 cm^{-1}) is predominantly occupied by absorption bands resulting from glucose, ethanol and glycerol, as expected due to the C-O stretching caused by their common alcoholic group. Furthermore, the phosphate group, too, exhibits dominant spectral absorbance in that region. The right-hand side of the fingerprint region (1350-1500 cm^{-1}) is predominantly occupied by absorption bands of acetic acid and ammonium and the peaks of both components appear rather separate. This is in line with the observation that these two molecules are composed of different functional groups compared to all the other components (carboxyl group of acetic acid and ammonium as an individual group itself) and thus result in a different spectral fingerprint. Regarding their molecular absorption, or respectively their IR activity, phosphate and glucose are at the very front with 3.5 and 3.3 Mol^{-1} followed by all the other components showing molar absorption between 2.8 Mol^{-1} (glycerol) and 2.5 Mol^{-1} (ammonium and acetic acid). This already suggests that phosphate and glucose might be more accurately predicted in lower concentrations compared to the other components due to their inherently higher molecular absorption.

Studying the molecular structure as well as the individual spectral absorption bands and molecular absorbance of each component led to valuable considerations: Based on common structural moieties, or functional groups, respectively, a certain overlapping of absorption bands could already be expected for glucose, ethanol and glycerol (common alcoholic group). Exhibiting a unique moiety compared to all the other analytes of interest (as *e.g.* acetic acid and ammonium) is likely to result in distinctive absorption bands separated from the other components but might also result in overlapping spectral features (as *e.g.* observed for phosphate). Pure component spectra should always be investigated to complement the theoretical structural considerations. Overall it can be concluded that spectral features were highly overlapping within this study and cannot be calibrated by univariate regression. Glucose and phosphate show the highest molecular absorbance and dominate the region 1000 cm^{-1} -1200 cm^{-1} . Hence, PLS modeling of ethanol and glycerol occupying the same spectral region but exhibiting lower molecular absorbance might be prone to indirect correlation with the dominating species phosphate and glucose in that region. However, this is to be

investigated in order to come to a final conclusion (see section 2.3.4.4 Investigation of indirect predictions).

Finally, considering the composition of the medium itself, next to water, peptides and amino acids as main ingredients of yeast extract and peptone, respectively, will cause the spectral fingerprint of the background matrix during the fermentation. Strong water absorption appears at ca. 1630 cm^{-1} due to O-H stretching as mentioned before. Peptide or amide vibrations involve C=O, C-N and N-H groups of the peptide bond, or amide group, respectively. Spectral absorption of peptides due to amide I, amide II and amide III structures characteristically appear in the region $1600 - 1700\text{ cm}^{-1}$ (amide I), $1500-1600\text{ cm}^{-1}$ (amide II) and $1200 - 1350\text{ cm}^{-1}$ (amide III), whereas the amide III signal is significantly weaker than the amide I and II signals [44]. As for the analytes of interest spectral characterization is found in the region $900-1550\text{ cm}^{-1}$, and only amide III signals might interfere. As amide III absorption bands are described to be rather weak and amide I and II absorption bands are characterized outside the spectral region under study, the YP background matrix was not expected to cause significant interferences. However, the acid used for pH control, H_2SO_4 , or the SO_4^{2-} ion, respectively, exhibits high IR-activity and causes a characteristic peak at 1100 cm^{-1} [16]. Thus, the addition of acid might interfere with the components glucose, ethanol, glycerol and phosphate. The variability introduced by the acid addition was not taken into account within this work. Acid was added automatically to control the pH in the second half of the fermentation. The first part of the fermentation is characterized by the production of acidic metabolites, and thus base addition is required to keep the pH at the operational set-point of 6. Hence, during the frequent sampling over the first 14-16 h of fermentation, acid addition was not a major concern. As mentioned before, the base used for pH control (KOH) was not expected to cause spectral variation other than a peak at around 1630 cm^{-1} due to the O-H stretch, which was located outside the spectral region of interest. Both, acid and base addition profiles can be reviewed in the supplementary material, Figure S2.1.

2.3.2.2 Fermentation spectra – raw spectra analysis

As general reference, the spectral evolution of data set 4 is shown (Figure 2.9) representing a standard batch with the possibility to compare cell free and cell-containing samples.

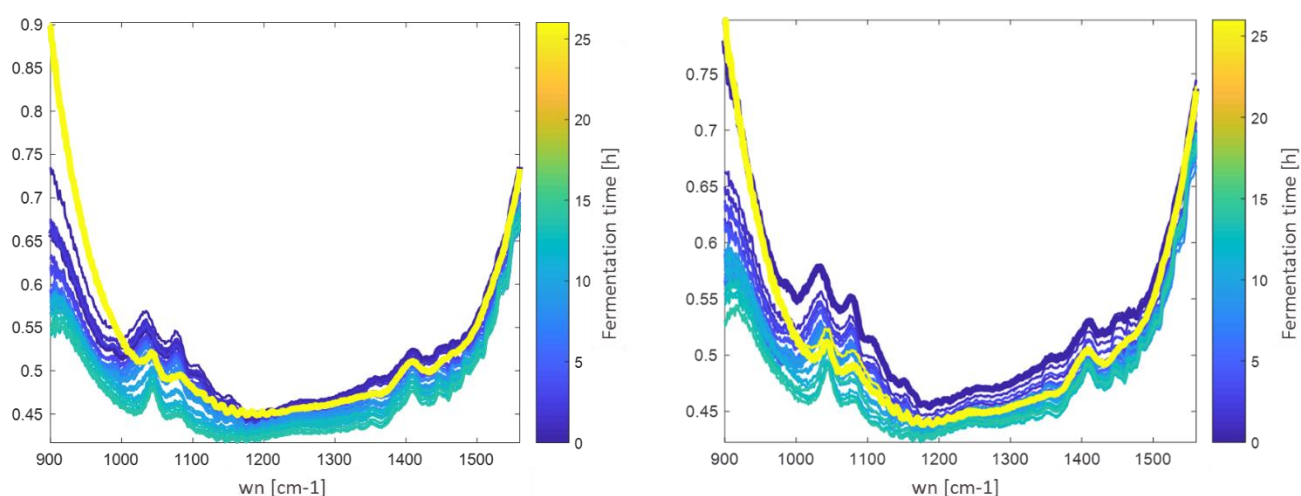


Figure 2.9: Spectral evolution of fermentation 4 (Table 2.1). Left: Cell free samples. Right: Cell-containing samples

Figure 2.9 was investigated with respect to: (1) the time trend observed of the spectra over the fermentation course; (2) the difference between cell free and cell containing samples and (3) the most dominant absorption bands. In both cases, a clear gradient can be observed from 0 to 14 h, when spectra were collected hourly. The last spectra, acquired after 26 h, takes a big 'jump'. When considering the last spectrum, the trend observed in the first spectra (0 - 14 h) seems to be reversed, but as no spectra in between 14 and 26 h were collected, this remains an assumption. However, as the time from 14 – 26 h is characterized by metabolite consumption, whereas the first 13 h are mainly defined by metabolite production, the assumption seems reasonable from a metabolic point of view. Furthermore, cell-free and cell-containing samples did not differ significantly from each other, which was also confirmed by PCA analysis of the raw spectra (Figure 2.10).

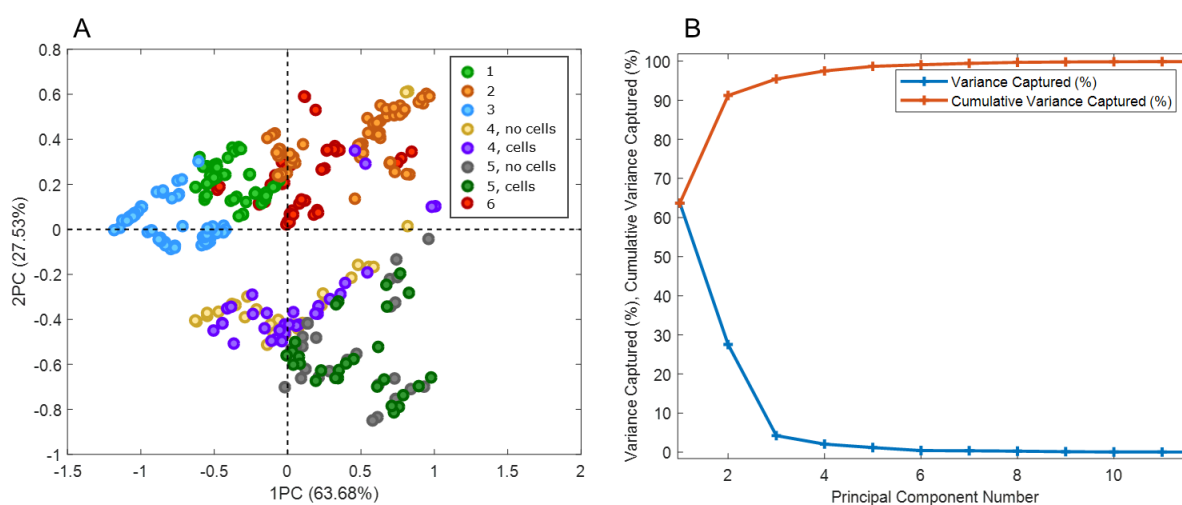


Figure 2.10: PCA of the raw spectra of data sets 1-6. A) Score plot for the first two principal components. Data set 4 and 5 were analysed with and without cells, as indicated in the legend. B) Explained variance per principal component (blue line) and cumulative variance explained (orange line).

Figure 2.10 A shows the scores of data sets 1 - 6 on the first two principal components. Data set 4 and 5 are clearly separated from the other data sets by the second principle component, which can be explained by a change in the optics of the NLIR instrument, which needed to be newly aligned in between. However, cell-free and cell containing samples of data set 4 and 5 lie mostly on top of each other and differences between cell-free and cell-containing samples are minor. Thus, it can be concluded that no significant spectral difference results from cells inside the sample (and that a consistent instrumental set-up is of course preferable to avoid differences between data sets due to the instrumentation itself).

According to Figure 2.10 B, about 95 % of the spectral variance is explained by the first 3 principal components, while after 6 principal components, explaining 99 % of the total variance, no further improvement of the PCA model was found by including additional principal components. The number of principal components, explaining most of the spectral variation, indicates the chemical rank of the spectra. In other words, the chemical rank reveals how many parameters (components) can be extracted and modeled independently from the IR measurements. Hence, based on Figure 2.10 B, it can be expected that 3 to 4 components are captured independently by a PLS model, while 2 – 3 components will be modeled based on indirect correlation to other components.

The main spectral absorption bands found in the fermentation samples shown in [Figure 2.9](#) occur in the region $1000\text{--}1200\text{ cm}^{-1}$ and $1350\text{--}1450\text{ cm}^{-1}$. In alignment with the raw spectra analysis of the pure components ([Figure 2.7](#) and [2.8 A](#)) the lower spectral region of $1000\text{--}1200\text{ cm}^{-1}$ represents the components glucose, glycerol and ethanol due to the C-O stretch of their common alcoholic group as well as phosphate, that was found to cause absorbance characteristics in the same region. Moreover, the spectral fingerprint of glucose ‘shines through’, dominating this lower region and spectral variation due to glycerol, ethanol, acetic acid (and phosphate) must represent fine differences in the dominating glucose absorption bands, not visible to the naked eye. The upper region $1350\text{--}1450\text{ cm}^{-1}$ represents the analytes ammonium and acetic acid. The complexity of spectral analysis is represented here: As the spectral fingerprint of two components in the same region is additive, only one dominating peak results in this upper region. Fine spectral nuances separating the 2 components ammonium and acetic acid are not visible to the naked eye and can only be captured by applying multivariate data analysis techniques such as PLS modeling. This holds as well for the other components glucose, ethanol, glycerol and phosphate, showing covariance structures in the mentioned lower region. As expected, significant absorption changes due to peptides, or, respectively, amide III absorption bands described in the region $1200\text{--}1350\text{ cm}^{-1}$, are not clearly visible in the spectrum.

2.3.4 PLS modelling

2.3.4.1 Exploring the data and preprocessing

PLS models were built on first derivative spectra (Savitzky-Golay (25, 3, 1) as mentioned in the materials and methods, section 2.2.4.1 PLS calibration and PCA analysis of spectral data. When inspecting the PCA model of the first derivative spectra, presented in [Figure 2.11](#), the benefit of this preprocessing, known to reduce noise and yielding a smoothening of the signal, becomes clear.

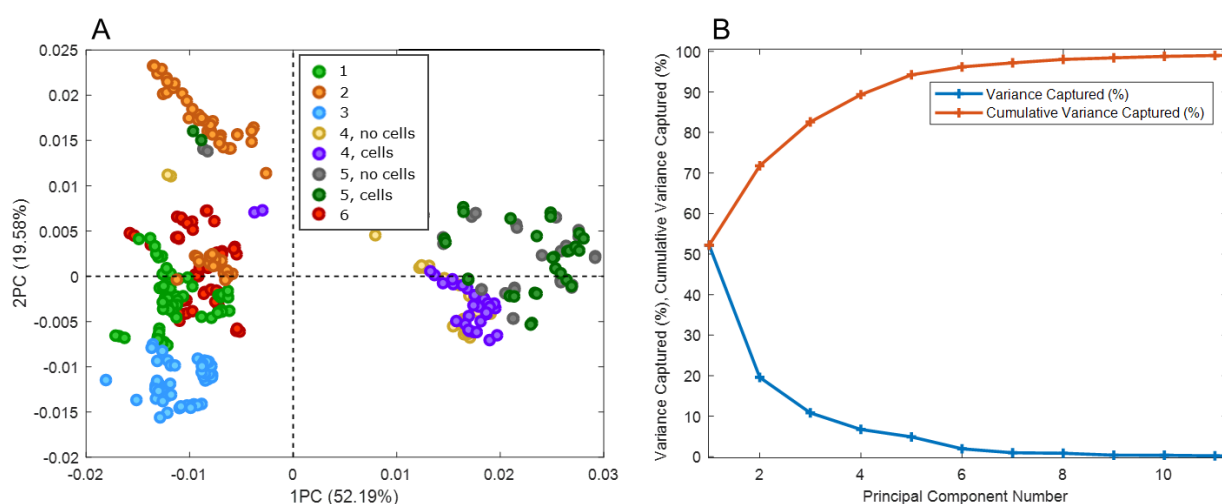


Figure 2.11: PCA of the first derivative spectra of data sets 1-6. A) Score plot for the first two principal components. Data set 4 and 5 were analysed with and without cells, as indicated in the legend. B) Explained variance per principal component and cumulative variance explained.

The score plot for the first two principal components of the first derivative spectra ([Figure 2.11 A](#)) looks similar to the one of the raw spectra presented in [Figure 2.10 A](#). The only difference between

the score plot of the raw spectra and the score plot of the first derivative spectra is the separation of data set 4 and 5. In [Figure 2.11 A](#), this happens along the first principal component explaining 52 % of the variance. In [Figure 2.10 A](#) (PCA model of the raw data), data set 4 and 5 were separated along the second principal component explaining 27 % of the variance. In other words, the arrangement in the optics of the NLIR instrument, to which the separation of data set 4 and 5 comes back, becomes even more evident in the preprocessed data. However, more important is the chemical rank of the first derivative spectra, that is clearly extended from 3-4 (raw spectra) to 6 in the first derivative spectra. Accordingly, the chance to model the 6 analytes of interest individually improves as a result of the preprocessing. Other preprocessing methods such as SNV and 2nd derivative were applied, however the overall best results were achieved with first derivative as preprocessing technique, clearly improving the modelling results.

2.3.4.2 Several PLS-models – general observations

PLS models were built on different calibration sets as outlined in [Table 2.3](#). Model 1 was chosen to show that using fermentation data only for PLS model calibration is not enough due to the highly correlated process dynamics of the analytes of interest. Model 2 was built to show that using synthetic samples only is not enough either, as the variation connected to the process background is missing. Models 3-8 were chosen to highlight different aspects of creating the ideal case of combining real and synthetic samples.

Table 2.3: Overview of calibration and validation sets used for PLS models 1 – 8. The data sets 1-6 are described in Table 2.1. For each model, the correlation coefficient for the calibration data (obtained as the average of all the different correlation coefficients between two components (neglecting positive or negative correlation), the total number of calibration samples and the concentration range (0 – max) per analyte in mM in both, the calibration and the validation data is given. Analytes in the last column are marked with a star when the validation range was exceeding the calibration range.

Model	Data sets used for Calibration	Number of samples for calibration	Average correlation coefficient in the calibration data	Data sets used for validation	Concentration Range Calibration / Validation [mM]
1	1	20	0.506	2, 3	Glucose*: 106 / 123 Glycerol*: 8 / 23 Acetic a.*: 16 / 26 Ethanol: 253 / 215 Ammonium*: 7 / 105 Phosphate*: 8 / 138
2	6	15	0.329	1, 2, 3, 4	Glucose: 144 / 130 Glycerol: 63 / 23 Acetic a.: 38 / 26 Ethanol*: 152 / 253 Ammonium*: 97 / 105 Phosphate*: 112 / 138
3	1, 6	35	0.293	2, 3, 4	Glucose: 144 / 130 Glycerol: 63 / 23 Acetic a.: 38 / 26 Ethanol: 253 / 215 Ammonium*: 97 / 105 Phosphate*: 112 / 138
3a	1, 6	35	0.293	4, 5	Glucose*: 144 / 226 Glycerol: 63 / 36 Acetic a.*: 38 / 43 Ethanol: 253 / 251 Ammonium*: 97 / 120 Phosphate*: 112 / 121
4	5, 6	31	0.177	1, 2, 3, 4	Glucose: 226 / 130 Glycerol: 63 / 23 Acetic a.: 43 / 26 Ethanol*: 251 / 253 Ammonium: 120 / 105 Phosphate*: 121 / 138

Table 2.3 continued

5	1, 2, 3, 6	53	0.250	4, 5	Glucose*: 144 / 226 Glycerol: 63 / 36 Acetic a.*: 38 / 43 Ethanol: 253 / 251 Ammonium*: 105 / 120 Phosphate: 138 / 121
6	1, 3, 6	54	0.273	2, 4, 5	Glucose*: 144 / 226 Glycerol: 63 / 36 Acetic a.*: 38 / 43 Ethanol: 253 / 251 Ammonium*: 97 / 120 Phosphate*: 112 / 138
7	3, 6, 4 (cell samples only)	50	0.265	1, 2, 5	Glucose*: 144 / 226 Glycerol: 63 / 36 Acetic a.*: 38 / 43 Ethanol: 253 / 251 Ammonium*: 97 / 120 Phosphate*: 112 / 138
8	3, 5, 6	66	0.230	1, 2, 4	Glucose: 226 / 130 Glycerol: 63 / 19 Acetic a.: 43 / 26 Ethanol*: 251 / 253 Ammonium: 120 / 105 Phosphate*: 121 / 138

Table 2.3 shows that, with the exception of model 1, built on fermentation data only, the average correlation coefficient between two components in the calibration set was between 0.18 and 0.33 due to the inclusion of synthetic and / or spiked samples (data set 5, 6). By incorporating spiked samples, the process dynamics were successfully decoupled. Furthermore, as the fermentation data sets showed a high batch-to-batch variability (Figure 2.5), it was not possible to completely match the calibration and validation space for all components, in any of the models. In each validation set, at least one component lies outside the calibration range. Components outside the calibration space are marked with a star in Table 2.3.

The performance of model 1-8 was evaluated based on the relative RMSEP obtained from the respective validation set. It must be emphasized that the performance of PLS calibration models was evaluated based on external validation, that is to say, data sets used for validation were not included in the calibration procedure. This generally gives a more realistic estimate of model performance than e.g. only inspecting the predictions based on cross validation (internal validation). The relative RMSEP values calculated over the entire samples used for validation (no differences between data sets are considered here) are summarized in Table 2.4. Model 3 and 3a are of course the same. They were divided into 3 and 3a to emphasize that the model performance depends on the data quality used for validation. Here, data the quality is described as the matching of calibration and validation space.

Table 2.4: RMSEP of Models 1-8 in %. Minimum and average values refer to the minimum error of prediction (marked in green) and the average error of prediction (column average) achieved in the 8 models. All values are given in percent. Model 3a (marked in yellow) is presented in detail in section 2.3.4.3 Detailed presentation of Model 3a.

Model	Glucose	Glycerol	Acetic Acid	Ethanol	Ammonium	Phosphate
1	22.4	39.7	29.4	30.8	58.7	55.4
2	12.3	163.7	55.7	22.7	14.8	18.8
3	9.9	108.3	43.8	13.2	9.5	11.9
3a	5.6	41.9	40.5	33.6	11.8	17.3
4	6.9	46.9	45.3	31.7	18.9	9.6
5	44.1	218.9	70.8	52.8	20.0	47.1
6	13.4	106.1	39.5	27.8	12.5	11.7
7	3.8	55.5	53.6	25.9	15.0	10.3
8	8.3	61.9	41.7	31	22.7	7.9
Average	15.1	100.1	47.5	29.5	21.5	21.6
Min	3.8	39.7	29.4	13.2	9.5	7.9

Generally, good results were obtained for glucose, ethanol, ammonium and phosphate with RMSEP values below 15 % in the best cases (marked green in Table 2.4). Models for glycerol and acetic acid showed RMSEP values of 40 and 30 %, respectively, and higher. In the best case, RMSEP values for glucose, ethanol, ammonium and phosphate equaled 3.8, 13.2, 9.5 and 7.9 %, respectively, as outlined in green in Table 2.4. With that, the models for these components would be totally satisfying for online data acquisition in an industrial production site. Glycerol and acetic acid are modeled poorly and reasons for that will be discussed in section 2.3.4.4 Investigation of indirect predictions. Comparing the performance of the different models and considering the matching / mismatching of the calibration and validation space, it becomes obvious that, in models 3-8, better or worse performance boiled essentially down to the number of samples being inside or outside the calibration range. Furthermore, a combination of one fermentation set and the synthetic samples representing 35 samples in total was sufficient to produce a good calibration model, as exemplary demonstrated with model 3. Neither, the addition of more fermentation data nor the inclusion of more spiked samples yielded a significant further improved of the overall model performance.

The importance of decoupling the process correlation and the matching of calibration and validation space were shown and found to be crucial aspects for successful model calibration and validation. Some adjustments would be considered if preparing these data sets again: In order to match validation and calibration space, each fermentation would be run as duplicate in order to provide a calibration and external validation set that are matching each other. Besides, when considering reasonable process variations, the fermentation data and spiked fermentation samples used for spectral and reference analysis would be collected first and synthetic samples would be designed based on the overall variation found in the data sets. Finally, the design of synthetic samples would be slightly optimized as follows. Instead of permuting the concentration values for each analyte based on a reference time profile, they would be permuted based on equally spaced concentration

values between the absolute minimum and maximum value found for each component, across the fermentations performed. Thus, time traces would not remain at all in the synthetic sample set. Moreover, equal spacing is described and was found (data not shown) to decrease the (average) leverage value per data point, or respectively, across the entire data set. The leverage is an indicator of predictability of the respective data point within the calibration space, ranging from 0 -1. The lower the leverage value, the higher the confidence in, or respectively, the predictability of the corresponding data point [45]. Finally, a theoretical trial of applying equal spacing between the minimal and maximal concentration value found for each component, with a total number of 15 samples, clearly showed to decrease the correlation coefficients between the components to a value of 0.19 and below (see supplementary material, Figure S2.2, Table S2.5). A value of 0.23 was obtained in the synthetic samples designed as described in section 2.2.1.2 Design of synthetic samples. A number of 15 synthetic samples was chosen since a fermentation batch profile was assembled well with 15 samples within this study, spanning the minimal to maximal concentration range for each component. In order to further increase the impact of the synthetic samples in the calibration set, the number of synthetic samples could be increased to 20 or above, depending on how many fermentation samples are considered to represent the calibration set. An increased number of synthetic samples is expected to decrease the correlation coefficients between the analytes and therefor also the leverage value, which is expected to result in more robust models.

2.3.4.3 Detailed presentation of Model 3a

As model 3 was identified as an overall satisfying model, built on calibration data spanning the minimum and maximum concentration well for most components (in particular phosphate exceeded the calibration space, see [Table 2.3](#)), this model will be presented and discussed in detail. Besides, in order to match validation and calibration space, model 3 was validated on data set 4 and 5, in two separate approaches, referred to as model 3a in [Table 2.3](#). In the first approach, the data sets 4 and 5 were evaluated as they are. In a second approach, outliers (only occurring in data set 5) were eliminated from the validation set. The reference values for data set 4 and 5 can be found in the supplementary material, Table S2.6. Outliers excluded in the second approach are marked in red. Thus, the chosen validation sets offered the possibility to discuss:

- 1) the performance on cell-free compared to cell-containing samples as a measure of the impact of cells
- 2) the performance on spiked compared to non-spiked samples as a measure of model robustness
- 3) the effect of concentration outliers on the model performance.

The PLS calibration models built on the calibration sets and applied to the validation sets according to model 3a and their performance are presented in [Figure 2.12 – 2.14](#). Each figure presents the model for two components and relevant model performance parameters. The division of the figures is done with respect to the components and each component is presented on one side of the figures (left-hand side and right-hand side). When comparing the fermentation profile predicted with the fermentation profile measured by HPLC and colorimetric acid kits, respectively, the confidence interval for each measurement point is given. In most cases, it is very small and cannot be

distinguished visually from the actual measurement trend. However, a small deviation can be observed for some data points and this is when a standard deviation of greater than 5 % was present.

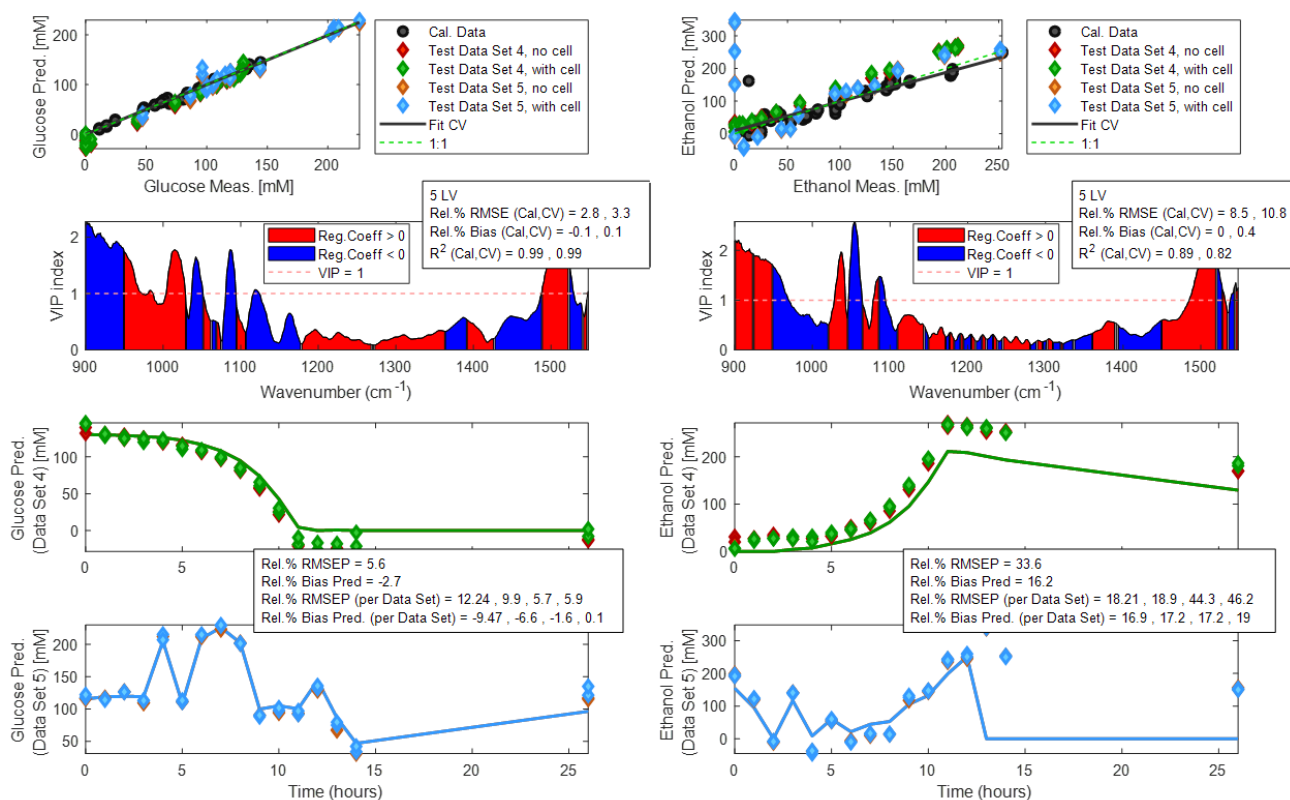


Figure 2.12: PLS calibration model and model performance for glucose (left) and ethanol (right). Calibration and validation sets correspond to model 3a in Table 2.3. For both models from top to bottom:

PLS calibration model;

VIP index and indication of positive (red) or negative (blue) regression coefficient as a function of the wavenumber. Additional model parameters relevant to the model calibration as outlined in the box: number of latent variables (LV), relative RMSE, bias and R^2 of calibration (Cal) and cross validation (CV) per data set 1 and 6, in this order;

Visual impression of model performance on data set 4: prediction (green and red diamonds as indicated in the legend (top figure) referring to cell-free and cell-containing samples) and actual sample concentration (solid line) in mM as a function of fermentation time in h. Additional information relevant to the model validation is outlined in the box: Relative RMSEP and bias calculated on the entire validation set comprised of the cell-free and cell-containing samples of data set 4 and 5. Relative RMSEP and bias per sub- data set in the order: set 4, cell-free samples; set 4, cell-containing samples; set 5, cell-free samples, set 5, cell-containing samples.

Visual impression of model performance on data set 5: prediction (blue and orange diamonds as indicated in the legend (top figure) referring to cell-free and cell-containing samples) and actual sample concentration (solid line) in mM as a function of the fermentation time in h;

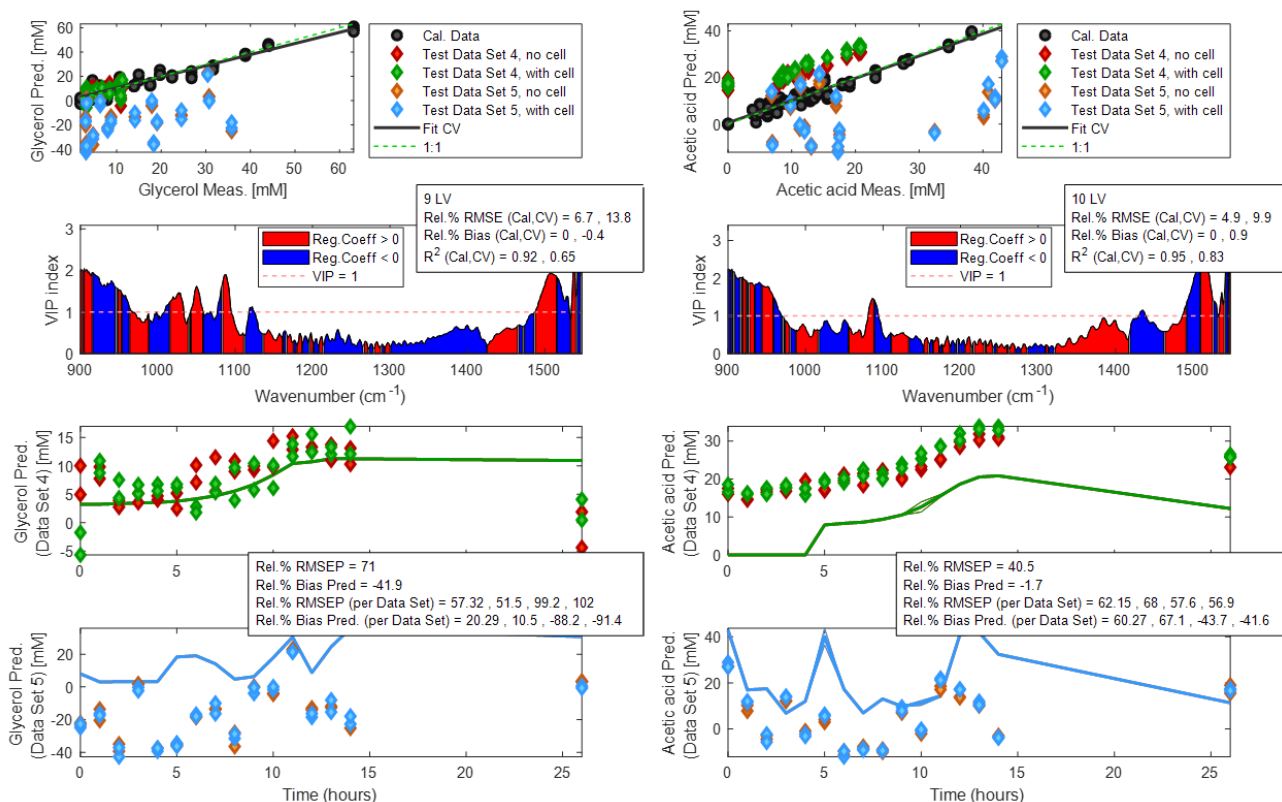


Figure 2.13: PLS calibration model and model performance for glycerol (left) and acetic acid (right). Calibration and validation sets correspond to model 3a in Table 2.3. For both components from top to bottom:

PLS calibration model;

VIP index and indication of positive (red) or negative (blue) regression coefficient as a function of the wavenumber; Additional model parameters relevant to the calibration set as outlined in the box: number of latent variables (LV), relative RMSE, bias and R² of calibration (Cal) and cross validation (CV) per data set 1 and 6, in this order;

Visual impression of model performance on data set 4: prediction (green and red diamonds as indicated in the legend (top figure) referring to cell-free and cell-containing samples) and actual sample concentration (solid line) in mM as a function of fermentation time in h. Additional information relevant to the model validation is outlined in the box: Relative RMSEP and bias calculated on the entire validation set comprised of the cell-free and cell-containing samples of data set 4 and 5. Relative RMSEP and bias per sub-data set in the order: set 4, cell-free samples; set 4, cell-containing samples; set 5, cell-free samples, set 5, cell-containing samples.

Visual impression of model performance on data set 5: prediction (blue and orange diamonds as indicated in the legend (top figure) referring to cell-free and cell-containing samples) and actual sample concentration (solid line) in mM as a function of the fermentation time in h.

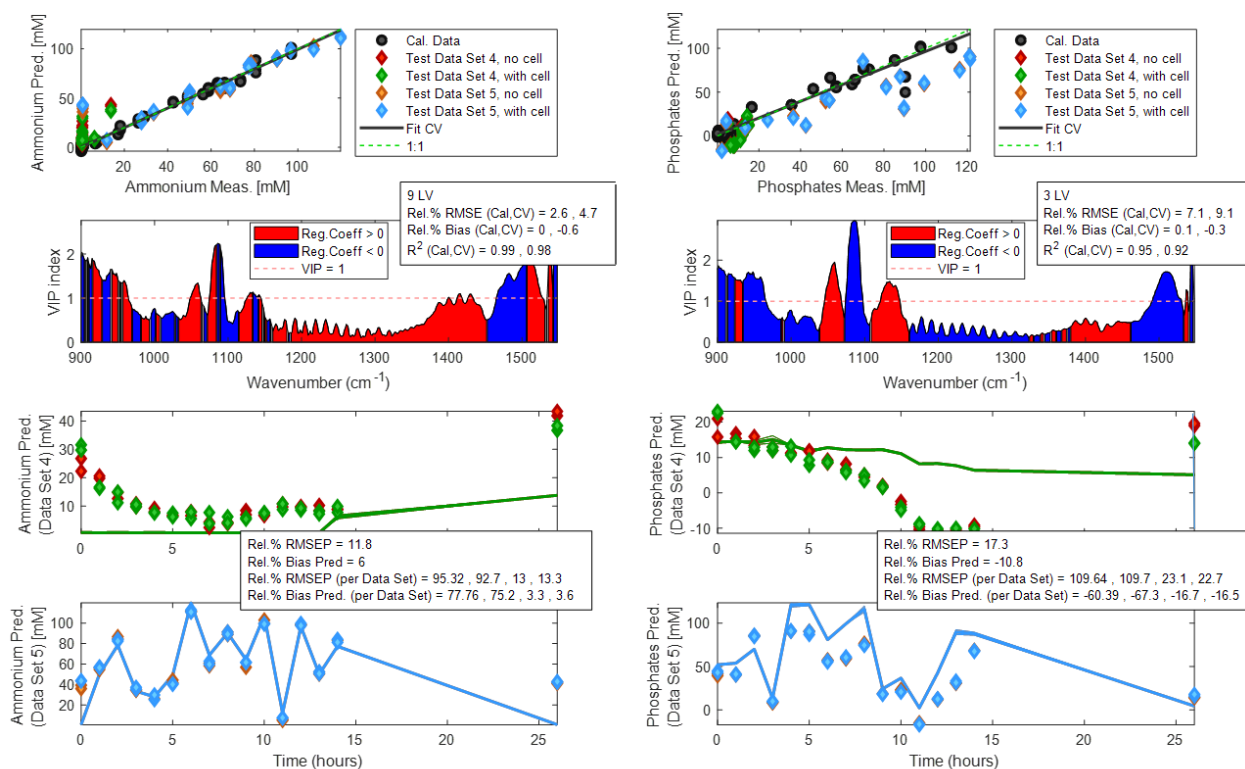


Figure 2.14: PLS calibration model and model performance for ammonium (left) and phosphate (right). Calibration and validation sets correspond to model 3a in Table 2.3. For both components from top to bottom:

PLS calibration model;

VIP index and indication of positive (red) or negative (blue) regression coefficient as a function of the wavenumber; Additional model parameters relevant to the calibration set as outlined in the box: number of latent variables (LV), relative RMSE, bias and R² of calibration (Cal) and cross validation (CV) per data set 1 and 6, in this order;

Visual impression of model performance on data set 4: prediction (green and red diamonds as indicated in the legend (top figure) referring to cell-free and cell-containing samples) and actual sample concentration (solid line) in mM as a function of fermentation time in h. Additional information relevant to the model validation is outlined in the box: Relative RMSEP and bias calculated on the entire validation set comprised of the cell-free and cell-containing samples of data set 4 and 5. Relative RMSEP and bias per sub-data set in the order: set 4, cell-free samples; set 4, cell-containing samples; set 5, cell-free samples, set 5, cell-containing samples.

Visual impression of model performance on data set 5: prediction (blue and orange diamonds as indicated in the legend (top figure) referring to cell-free and cell-containing samples) and actual sample concentration (solid line) in mM as a function of the fermentation time in h.

The obtained calibration model, typical model parameters linked to the calibration process, VIP scores and regression coefficient, and the visualization of the model performance per data set 4 and 5 are presented for each model. A VIP (very important variable) index greater than one indicates high importance of the respective variables (wavenumbers) for the model. A positive regression coefficient (red in Figure 2.12 – 2.14) indicates a positive correlation with the spectrum ('the higher the concentration the higher the absorption peak'), a negative regression coefficient (blue in Figure 2.12 – 2.14) indicates a negative correlation with the spectrum ('the higher the concentration

the lower the absorption peak'). It must be taken into account that the raw data was subject to preprocessing applying the first derivative, thus resulting in a split of each positive peak (raw spectrum) into a positive and negative part (first derivative spectrum). Hence, only where a blue and a red section are not adjacent, a true negative correlation is found in the model. A negative correlation can for example be found for the regression coefficients for phosphate (Figure 2.14) in the region between 1200 and 1300 cm^{-1} . Furthermore, the number of latent variables (LV) together with the relative RMSE, bias and R^2 of calibration (Cal) and cross validation (CV) are outlined. The number of latent variables was decided based on the plot of RMSE(CV) vs the number of latent variables where a local or global minimum was observed. The number of LV reveals to a certain extent the model complexity and the model robustness. A robust model is usually built on a small number of latent variables while a high number of latent variables demonstrates a high model complexity, usually involving a lack of robustness. Model validation is presented as the overall relative RMSEP and bias (comprising all samples from the validation sets 4 and 5) as well as the RMSEP and bias per sub-data set. The sub-sets, respectively, refer to validation set 4 and 5 with respect to cell-free and cell-containing samples, in this order. The model parameters linked to the calibration and validation process presented as part of the Figures 2.12-2.14 are summarized in Table 2.5.

Table 2.5: Summary of the model parameters presented as part of the Figures 2.12-2.14.

Parameter	Glucose	Glycerol	Acetic Acid	Ethanol	Ammonium	Phosphate
Calibration (data set 1 and 6)						
No. of LV	5	9	10	5	9	3
R^2(Cal / CV)	0.99 / 0.99	0.92 / 0.65	0.95 / 0.83	0.89 / 0.82	0.99 / 0.98	0.95 / 0.92
RMSE [%] (Cal / CV)	2.8 / 3.3	6.7 / 13.8	4.9 / 9.9	8.5 / 10.5	2.6 / 4.7	7.1 / 9.1
Bias [%] (Calibration)	-0.1 / 0.1	0 / -0.4	0 / 0.9	0 / 0.4	0 / -0.6	0.1 / -0.3
Validation whole data set 4 and 5						
RMSEP [%]	5.6	71	40.5	33.6	11.8	17.3
without outliers	9.3	-	35.6	-	15	-
Bias [%] (Prediction)	-2.7	-41.9	-1.7	16.2	6	-10.8
without outliers	-5	-	6.5	-	8.5	-

Table 2.5 continued

RMSEP [%] per individual subset in data set 4 and 5 (cell-free (-) / cell-containing (+))						
Data set 4 (- / +)	12.2 / 9.9	57.3 / 51.5	62.15 / 68	18.2 / 18.9	95.3 / 92.7	109.6 / 109.7
Data set 5 (- / +)	5.7 / 5.9	99.9 / 102	57.6 / 56.9	44.3 / 46.2	13 / 13.3	23.1 / 22.7
Without outliers	11.8 / 12.2	-	65.7 / 49	-	17.1 / 17.3	-
Prediction Bias [%] (cell-free (-) / cell-containing (+))						
Data set 4 (- / +)	-9.5 / -6.6	20.3 / 10.5	60.3 / 67.1	16.9 / 17.2	77.8 / 75.2	-60.4 / -67.3
Data set 5 (- / +)	-1.6 / 0.1	-88.2 / -91.4	-43.7 / -41.6	17.2 / 19	3.3 / 3.6	-16.7 / -16.5
Without outliers	-4.3 / -1.1	-	-43.9 / 32.3	-	5.9 / 6.2	-

Firstly, with respect to [Table 2.5](#) and [Figures 2.12-2.14](#), the yielded calibration is discussed. The calibration curve of each component, as presented in the upper section of the [Figures 2.12-2.14](#), resulted in fairly good R^2 values ranging from 0.99 (glucose) to 0.89 (ethanol). Ethanol showed a clear outlier in the calibration curve ([Figure 2.12](#), right) and the reduced R^2 value compared to the other calibration curves, reaching a R^2 value higher than 0.92, must be related to that. The good R^2 values of the different calibration curves was in alignment with an overall small bias in calibration and cross-validation (lower than 1%). However, it can be observed that the calibration data could be improved. Especially glycerol and acetic acid are calibrated in an extended range that was not relevant to the fermentation course (data set 4). Instead of extending the calibration range, an equally spaced calibration set with a reasonable number of samples in the range actually occurring in the fermentation is expected to improve the model. The RMSE of calibration was overall below 10 %, the RMSE of cross-validation overall below 15 % suggesting fairly good predictions for all components. The difference between RMSE of calibration and cross-validation is biggest for the components glycerol and acetic acid (7 and 5 %) indicating challenged predictions for these components. The number of latent variables (LV) upon which the models were built suggested good model robustness (future predictability) for phosphate (3 LV), glucose and ethanol (5 LV each). Models for glycerol, acetic acid and ammonium were built upon 10 and 9 LV, a rather high number, suggesting that these models are less robust and likely dependent on covariance structures, which must be conserved for achieving reliable future predictions.

Secondly, the evaluation of model performance with respect to the validation data sets 4 and 5 is discussed. The model performance was evaluated according to the three main points of interest as presented in detail in section 2.3.4.3 Detailed presentation of Model 3a. Note that, the assumption that glycerol and acetic acid are modeled based on indirect relationships is proven in section 2.3.4.4 Investigation of indirect predictions. However, in the following section, this is expressed as a hypothesis.

2.3.4.3.1 The impact of the presence of cells on the model performance

In alignment with the PCA data presented in Figure 2.11, the presence of cells did not affect the prediction results significantly. This was confirmed in Figures 2.12-2.14 (lower half of the figures) and is in accordance with literature, describing that cells can only be predicted indirectly due to the very low penetration depth of ATR-MIRS [10]. In Figures 2.12 – 2.14, the predictions of cell-free and cell-containing samples lie essentially on top of each other, in both natural (data set 4) and spiked (data set 5) fermentation samples. However, a small, yet clear, off-set between cell-free and cell-containing samples becomes obvious in case of the models for glycerol and acetic acid. These two components showed consistently the highest prediction error across the models built, in this set-up (model 3a) of 71 % and 40.1 %, respectively. It suggests that cells might slightly alter parts of the spectrum used for the prediction of these two components. Since their prediction was challenged across all models built, cells seem to affect the prediction of these two components somewhat more than the prediction of the components glucose, ethanol, ammonium and phosphate. According to literature, yeast cells cause marginal absorption bands at 1239 cm^{-1} (P=O stretch of phosphodiester groups as e.g. found in the phospholipid membrane (cell –surface)) and around 1540 cm^{-1} (N-H stretch of e.g. membrane bound proteins) [46]. Figure 2.9 indeed shows a small peak around 1240 cm^{-1} in the spectra of cell containing samples of data set 4, which is essentially absent in the spectra of the cell-free samples. However, it also appeared during the early stages of the fermentation process in the cell-free samples. That suggests that this response might also result from the presence of peptides in the fermentation broth, expected to cause marginal spectral variation between $1200\text{--}1350\text{ cm}^{-1}$ due to amide III absorbance [44]. Finally, the respective peak at 1240 cm^{-1} might be an effect of both, peptides in the fermentation broth and cells, considering that cells do have peptides incorporated in the membrane. This would explain while the peak vanished in the cell-free samples but remained in the cell-containing samples. In this case, the peak was subject to both, peptide degradation and consumption, as well as microbial growth, respectively. Contrarily, the region of 1240 cm^{-1} was not of importance for the glycerol and acetic acid model (VIP scores smaller than 1 in Figure 2.13) while the upper region of 1540 cm^{-1} (eventually linked to cells) revealed a narrow range of high VIP scores for both analyst. Spectral variation at 1540 cm^{-1} due to cells was not visible to the naked eye in Figure 2.9 but the region was important to the modeling of glycerol and acetic acid according to the VIP scores (Figure 2.13). Thus, if cells caused spectral variation in this region, the prediction of glycerol and acetic acid would be biased.

It can be concluded that cells caused marginal spectral variation visible in the region around 1240 cm^{-1} effecting in particular the prediction of glycerol and acetic acid. Both components are generally modelled weak and likely based on indirect relationships. Thus, minor variation in regions of covariance caused by cells, as e.g. the upper spectral region around 1540 cm^{-1} , resulted in biased predictions for these two components. The models for glucose, ethanol, ammonium and phosphate were not affected by cells.

2.3.4.3.2 Model robustness

If a model responds correctly to spiking, altering the process dynamics and thus decoupling the correlations between the components, it can be related to a high model robustness (within process conditions calibrated). The models for glucose and ammonium, and to a fairly good extent the models for ethanol and phosphate, respond accurately to the spiking. This furthermore reveals that

in particular the models for glucose and ammonium, and to a less, yet still profound extend for ethanol and phosphate, are based on direct spectral correlation with distinct spectral features. This is confirmed by the VIP index for the 4 components, showing high values (higher than 1) for different peaks (peak-sections) in the spectrum (Figure 2.12 and 2.14). The actual fermentation course (data set 4) predicted for glucose and ethanol is in alignment with the reference values while the prediction of ammonium and phosphate is off. This is explained by the very low concentration values for both components in the fermentation broth, with only minor variation. The model was constructed for a reasonable variation of ammonium and phosphate concentrations within a range of 10 – 100 mM, while both components occurred in a concentration of less than 13 mM during the fermentation. Within this range, the model is poorly calibrated and thus, a reasonable prediction cannot be expected. Glycerol and acetic acid were poorly predicted in natural and spiked fermentation samples. The general trend was captured well in the natural samples (data set 4), while the model responded poorly to spiking. This indicates, once again, that the model might be built upon spectral covariance structures with other components or, in other words, based on indirect relationships. This is coherent with rather small VIP scores which seem to overlap especially with the VIP scores from phosphate and ammonium in the spectral region around 1100 cm⁻¹.

It can be concluded that robust models were yielded in particular for glucose and ammonium, followed by ethanol and phosphate, accurately responding to spiking. The robustness of the models for ammonium and phosphate was restricted to higher concentration levels as a result of the calibration range chosen (adapted from the previous work (Appendix 1)). Models for glycerol and acetic acid responded poorly to spiking, revealing a lack of model robustness, suggesting indirect predictions.

2.3.4.3.3 The effect of concentration outliers

Concentration outliers were identified to be present only in data set 5, consisting of 16 samples per component. They can be found as marked in red in Table S2.6 in the supplementary material. Four outliers for glucose, four for acetic acid and two for ammonium were identified. These samples were excluded from the validation set in the subsequent approach. Thus, the predictions in data set 5 were expected to improve for these components. Interestingly, this was only found to be the case for the prediction of acetic acid, in the case of cell-containing samples, consequently also improving the overall RMSEP of data set 4 and 5 (see Table 2.5). In the case of glucose and phosphate, the opposite was observed. The predictions changed for the worse. This can be explained by equation (2.5). The RMSEP was calculated with respect to the concentration difference between maximal and minimal value found (Δy_{val}) in the validation set per component. Excluding the concentration outliers from the glucose and phosphate data significantly reduced Δy_{val} resulting in an overall higher error of prediction. That also points out that a few data points exceeding the calibration range provoke rather good prediction errors by increasing Δy_{val} . It also reveals that the decision on how to calculate the prediction error is not easy when calibration and validation space are partly mismatching. It suggests that in such a case, the average or mean concentration value might give a more realistic estimate of the RMSEP.

Finally, as expected, Table 2.5 reinforces the general statement that model performance evaluated based on the RMSE of cross validation only gives an overoptimistic idea of the model performance on future samples. Using an external validation set as done for models 1-8, exemplarily demonstrated

on model 3a, the RMSE of cross-validation was consistently found to be lower than the RMSE of prediction.

2.3.4.4 Investigation of indirect predictions

As a general observation, modelling of glycerol and acetic acid resulted in poor prediction, no matter how well calibration and validation space were matching. The following six considerations point out, that these two components are difficult to model and their modelling might rely on indirect correlation to one or several other species:

- 1) Both components are present in small amounts (up to 25 mM only, [Figure 2.5](#)).
- 2) Both components show a rather low molecular absorption ([Figure 2.8 B](#)).
- 3) Process correlations of both components were decoupled. However, both components show overlapping spectral features with other components (glycerol with glucose, ethanol, and phosphate; acetic acid with ammonium ([Figure 2.8 A](#))).
- 4) Best predictions were achieved with model 1 (calibration based on fermentation data only, process correlations were not decoupled), showing the highest correlation coefficients in the calibration data employed ([Table 2.3 and 2.4](#)).
- 5) Models for both components poorly respond to spiking ([Figure 2.13](#)).
- 6) Models for both components are slightly affected by the presence of cells, causing variations in the spectral upper region (1540 cm^{-1}) where no spectral correlation ([Figure 2.8 A](#)) but high VIP scores for both models are observed ([Figure 2.13](#)).

Monitoring of glycerol and acetic acid by FT-IR spectroscopy during a yeast fermentation process was also found to be challenging by Schenk and coworkers [38]. This work reports comparable high prediction errors of 50 – 200 % for glycerol and acetic acid in uncorrelated samples and suggests a lower detection limit of 0.5 g/L for each component (corresponding to 5 mM glycerol and 8 mM acetic acid, respectively). A considerable number of samples below this suggested detection limit were present in the data sets employed in this work. The weak IR absorbance of both components and the low levels in which they are produced during the fermentation process definitely challenged the modelling process.

In order to investigate the assumption, that glycerol and acetic acid were poorly predicted since their predictions were based on indirect correlations, the correlation coefficient and VIP scores as a function of wave number was studied in detail for each component. This investigation was based on the preprocessed (first derivative) data for model 3. In such a way, the spectral fingerprint of each component (correlation coefficient with the spectrum) and the regions relevant to the model (VIP scores) could be visually compared for all components at once. [Figure 2.15](#) presents the average raw spectrum and the average preprocessed spectrum ([Figure 2.15 A](#)), a heat map of the correlation coefficients ([Figure 2.15 B](#)) and a heat map of the VIP scores ([Figure 2.15 C](#)). In [Figure 2.15 A](#), the effect of applying the first derivative on the raw spectrum is shown. Each peak was split into a positive and a negative variation. As a result of taking the first derivative, the spectral variation ('resolution') is enhanced but spectral interpretation becomes more difficult. This becomes obvious when looking at the correlation coefficients ([Figure 2.15 B](#)). If the raw spectra were the reference for the correlation

coefficients with the wave numbers, a positive correlation of a component and a certain wavenumber would result in a positive correlation coefficient. This positive correlation coefficient would indicate a direct correlation with the respective component. Due to the introduction for the derivative of the spectra, every peak was split in a positive and a negative counterpart, the latter resulting in a negative correlation coefficient. When based on the raw spectra, a negative correlation coefficient would indicate a negative correlation between the respective component and the spectra. This interpretation does not hold when inspecting the correlation coefficients of the first derivative spectra, as the negative correlation coefficient can be seen as an 'artefact' resulting from the preprocessing. With this in mind, the correlation coefficients (Figure 2.15 B) can be studied. Generally speaking, the correlation coefficients complement the findings made on the pure component and raw fermentation spectra analysis (Figure 2.8 A and 2.9). Essential absorption bands in the raw spectra are represented in an enhanced way in the preprocessed spectra, resulting in high correlation coefficients for the marked regions of major variation: The region from ca. 1000 – 1200 cm^{-1} (first inlet, 'lower spectral region') was demonstrated to represent evident absorption bands of glucose, glycerol, ethanol and phosphate (Figure 2.8 A). The region from ca. 1350 – 1500 cm^{-1} (second inlet, 'upper spectral region') showed evident absorption bands caused by the components ammonium and acetic acid. However, it was not obvious to the naked eye in the pure spectra that ammonium also shows correlation with the lower spectral region and that phosphate and ethanol also show correlation with the upper spectral region, as can be seen in Figure 2.15 B. Interestingly, high VIP scores (greater than 1), representing high relevance for the model, are only found for the lower spectral region, not in the upper spectral region (Figure 2.15 C). As both, high correlation coefficients and VIP scores of all components are broadly overlapping in the spectral borders and the lower spectral region, no explicit answer can be given to the assumption, that glycerol and acetic acid were modeled based on indirect relationships. What can be concluded, yet, is, that all components are modeled based on the same spectral regions, again confirming the assumption, that not all components are modeled independently from each other.

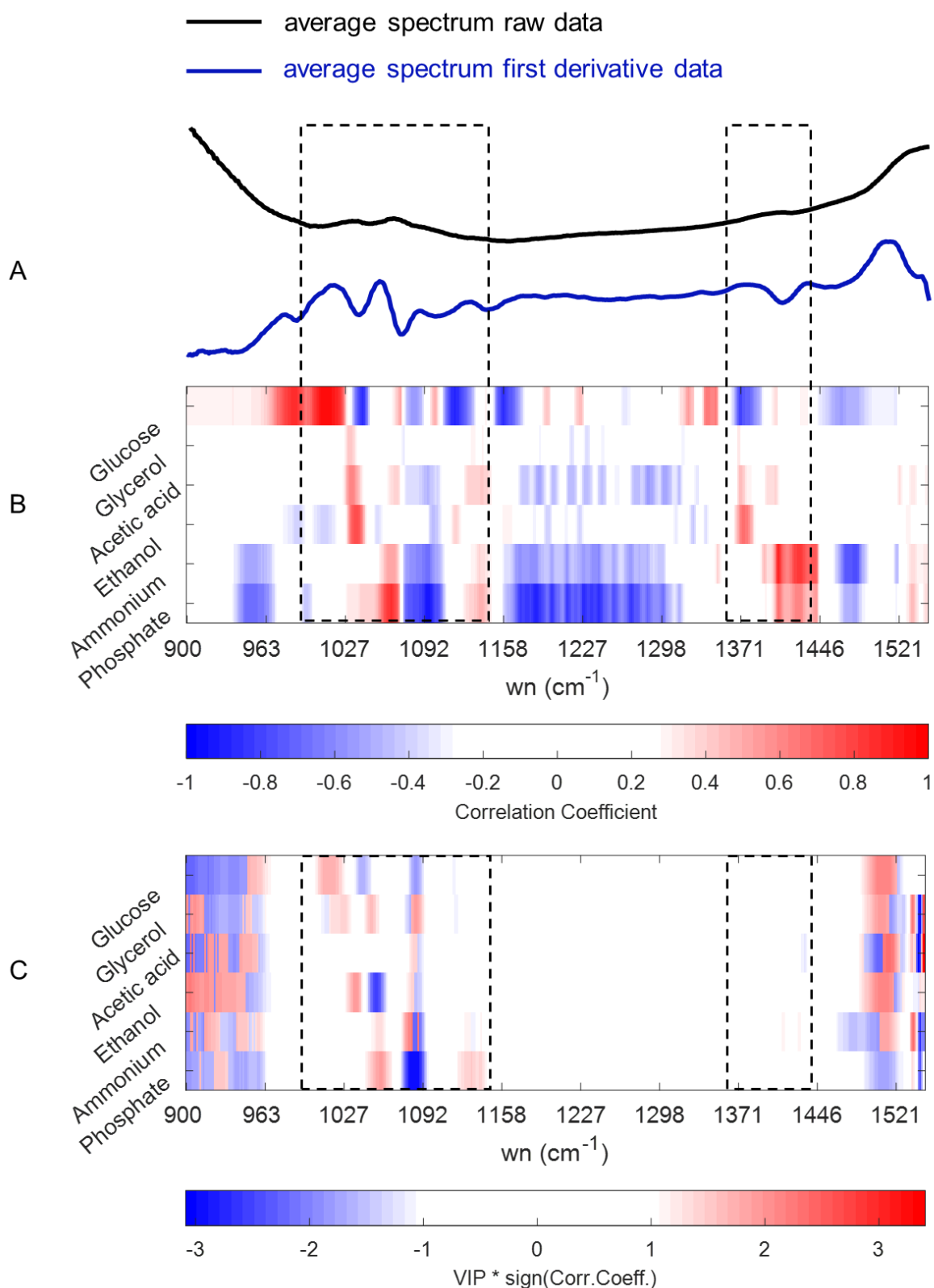


Figure 2.15: A) Averaged raw spectrum and averaged pre-processed spectrum of the calibration set for model 3. B) Heat map of correlation coefficients. C) Heat map of VIP scores. The inserts indicate spectral areas with evident absorption bands in the spectrum (compare to Figure 2.8 A).

Nonetheless, the investigation of the VIP scores proposes that a variable selection might be beneficial to the model. In such a case, [Figure 2.15 C](#) clearly advocates that the variables in the 965 – 1000 cm^{-1} region and the 1150 – 1380 cm^{-1} region did not contribute to the model, and thus might be excluded for the better. However, they might contribute to model stabilization and a variable selection as proposed must be performed in order to provide a prove of the hypothesis.

The assumption of indirect prediction based on overlaying spectral features and VIP scores was investigated further. As described in [24], two analytes are predicted by the same linear combination of the predictor data, if the predictions of these two analytes are correlated, regardless of how they correlate in the raw data. For simplicity, let's think about a components *a*, that would be glycerol or acetic acid in this case, and a components *b*, that would be glucose, ethanol, ammonium or phosphate. If the correlation between predicted *a* and predicted *b* is higher than the correlation between measured *a* and predicted *b*, the prediction of *a* is based on the same spectral variation as *b*. Hence, the prediction of *a* is indirectly correlated with variation associated with *b*. In other words, *a* is not predicted independently but based on indirect correlation with *b*. *Vice versa*, if the correlation between predicted *a* and predicted *b* is the same or very similar to the correlation between measured *a* and predicted *b*, *a* and *b* are predicted independently from each other. Their predictions are then based on distinct spectral features which PLS regression transformed into independent latent variables, associated with each component. Although particular focus lay on the investigation of the prediction of acetic acid and glycerol, all six analytes were investigated with respect to the occurrence of indirect covariance structures to one another. The outcome is presented in [Figure 2.16 A-F](#), investigating indirect predictions for the six analytes based on glucose (A), glycerol (B), acetic acid (C), ethanol (D), ammonium (E) and phosphate (F). The y-axis, exemplarily explained for [Figure 2.16 A](#), represents the correlation between predicted analyte concentration and measured glucose concentration (filled triangle) and, respectively, predicted analyte concentration and predicted glucose concentration (open square). The x-axis represents the model performance expressed in the correlation coefficient (R^2) between measured and predicted analyte. The [Figures 2.16 \(B\) – \(F\)](#) are to be read accordingly. Indirect predictions are indicated by a correlation between two analytes being higher for 'predicted vs predicted' than for 'predicted vs measured'. Practically speaking, to spot an indirect relation in [Figure 2.16](#), the filled triangle should show a significantly higher value on the y-axis than the corresponding open square, to which it is connected by a solid line. First of all, the prediction of glycerol was investigated. For that component, it was checked if its prediction was based on the prediction of one of the other components ([Figures 2.16 A, C, D, E, F](#)). It can be concluded that the prediction of glycerol is essentially based on a correlation with the prediction of ammonium, and to a smaller extent a correlation with phosphate and acetic acid. Comparing these result with the pure component spectra ([Figure 2.8 A](#)), correlation to acetic acid and phosphate can be expected (overlapping absorption bands in the region 1200 – 1300 cm^{-1}). Contrary to that, a correlation between glycerol and ammonium, explaining most part of the glycerol prediction (highest correlation coefficient on the y-axis compared to phosphate and acetic acid) is not as evident and rather surprising.

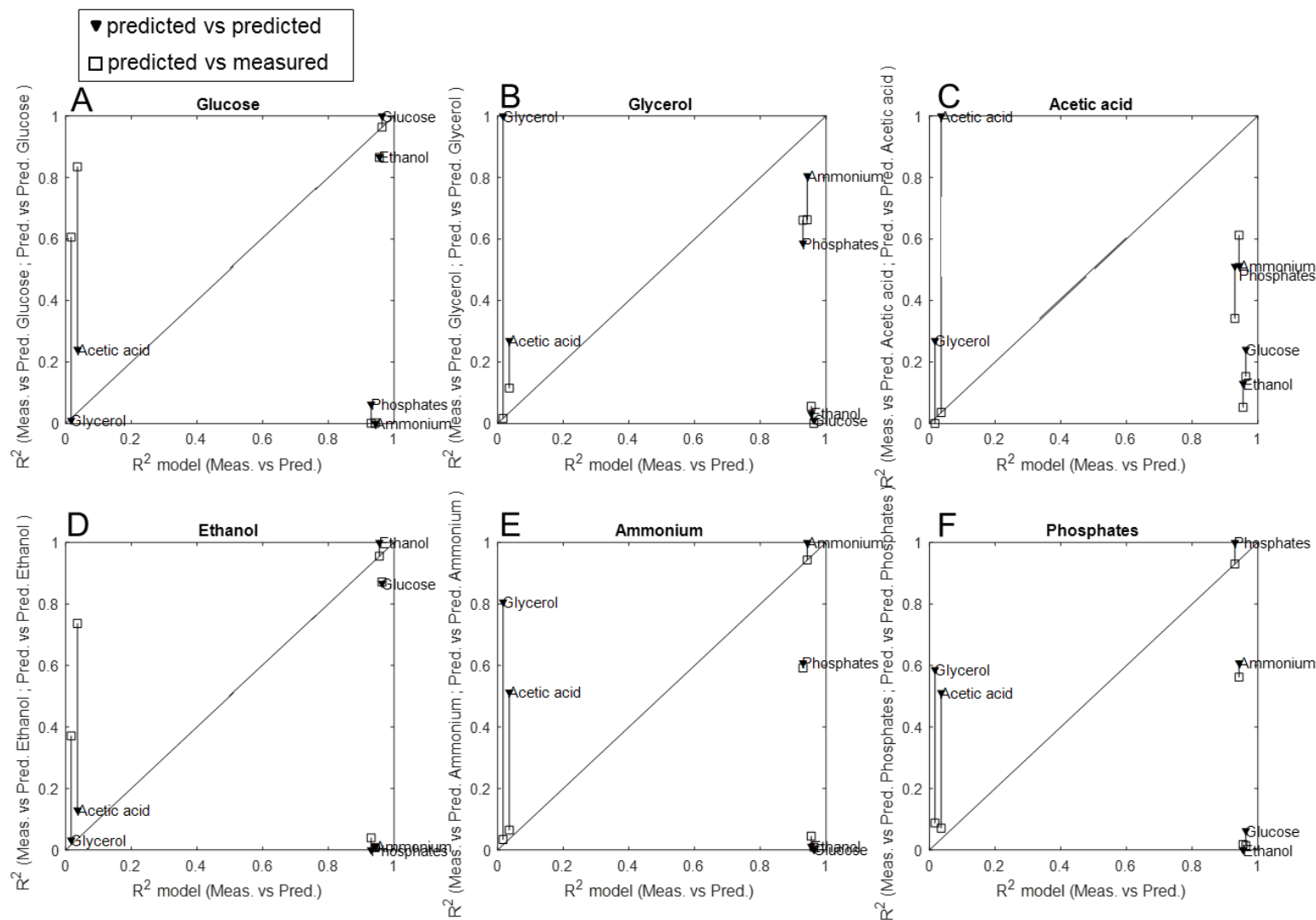


Figure 2.16: Investigation of indirect predictions based on A) Glucose, B) Glycerol, C) Acetic Acid, D) Ethanol, E) Ammonium and F) Phosphate. The y-axis, exemplarily explained for Figure 2.16 A), represents the correlation between predicted analyte concentration and measured glucose concentration (filled triangle) and, respectively, predicted analyte concentration and predicted glucose concentration (open square). The x-axis represents the model performance expressed by the correlation coefficient (R^2) between measured and predicted analyte.

Secondly, the prediction of acetic acid was investigated (Figure 2.16 A, B, D, E and F). It was found that acetic acid is equally explained by the prediction of ammonium and phosphate, and to a smaller extent by the prediction of glycerol. This is coherent with the pure component spectra, showing overlapping spectral features in the region $1250 - 1450\text{ cm}^{-1}$ (Figure 2.8 A). Furthermore, according to Figure 2.16 B and C, glycerol and acetic acid are to the same extent predicted by respectively the other component. This explains the covariance found between glycerol and ammonium: Glycerol is to a certain extent explained by cross-correlation to acetic acid, which in turn, is explained by ammonium and phosphate. *Vice versa*, it can be observed that the prediction of ammonium relies to a small extent on the prediction of glycerol (Figure 2.16 B) and the prediction of phosphate relies to a small extent on the prediction of acetic acid (Figure 2.16 C). Thirdly, it can be observed that the prediction of glucose and ethanol are independent from the prediction of any other component studied.

To conclude the findings based on Figure 2.16, within the current study, glucose and ethanol were being predicted independently while the predictions of glycerol and acetic acid was trapped in a cage of covariance with ammonium and phosphate. Thus, glycerol and acetic acid were not predicted independently but based on indirect correlation with each other as well as with ammonium and phosphate. While the cross-correlation of acetic acid, ammonium and phosphate resulted from overlapping spectral features (Figure 2.8 A), the covariance of glycerol and phosphate was linked by spectral covariance of glycerol to acetic acid. This explains why glycerol was subject to the overall poorest prediction within this study. The prediction of ammonium and phosphate, *vice versa*, were found to be somewhat related to the prediction of glycerol and acetic acid but essentially, can be considered as independent.

2.3.5 The effect of instrumentation (spectral quality) on the model performance

Within this study, models presented and discussed until now were built on spectral data acquired with the NLIR instrument, representing a prototype of its kind applying novel technology. In order to compare the spectral quality achieved with the NLIR instrument and a standard FT-IR instrument, model 3 was built and validated on both, data acquired with the NLIR instrument and data acquired with the standard FT-IR instrument. Model 3 was validated on data sets 2 and 3, using NLIR and FT-IR data, respectively. Spectra used were subject to the same preprocessing (first derivative) and the same spectral region ($900 - 1550\text{ cm}^{-1}$) was considered for evaluation. In this way, the model performance of each component, expressed as the relative RMSEP for both spectral data sets, could be interpreted as a measure of quality of the spectral data. The results are presented in Table 2.6.

Table 2.6: RMSEP in percent for all components with respect to model 3, evaluated on data set 2 and 3. Model 3 was built and evaluated with respect to spectral data acquired with the NLIR and the FT-IR instrument, respectively, as outlined in the table. The average value was calculated as the raw average, disregarding glycerol and acetic acid.

	Glucose	Glycerol	Acetic Acid	Ethanol	Ammonium	Phosphate
FT-IR	3.4 %	56.9 %	19.1 %	5.2 %	12.5 %	13.1 %
Average	8.6 %					
NLIR	10 %	133.9 %	43.1 %	8.7 %	7.4 %	14 %
Average	10.0 %					

As demonstrated in [Table 2.6](#), the model error is overall lower for data acquired with the standard FT-IR instrument, except for the component ammonium. This indicates a better sensitivity or, respectively, a better quality of the spectral data acquired with the FT-IR instrument (as expected). However, disregarding the models built for glycerol and acetic acid as found to be generally challenging, the difference can be considered as insignificant. Excluding glycerol and acetic acid from the comparison, on average, a model error of 8.6 % and 10 % was achieved with the standard FT-IR instrument and the NLIR instrument, respectively. Generally, with both instruments, models yielded an error of less than 15 %. It is important to notice that the change in spectral quality is not linear over the spectrum. The spectral quality is downgraded in most regions to different extents, as indicated by a higher RMSEP of models built on NLIR data. The difference in spectral quality is evident in the spectra themselves, as demonstrated in the supplementary material, Figure S2.6. For the components glucose, ethanol and phosphate the error changed for worse by 3.5 % on average. Other regions benefitted from the change in quality, as observed for the modeling of ammonium. In case of ammonium, the RMSEP value improved by 5 % when built on NLIR-data.

It can be concluded that the spectral quality is overall better with the FT-IR instrument directly influencing the model, as reflected in an overall lower model error (disregarding glycerol and acetic acid). However, the spectra acquired with the NLIR instrument are totally suitable for quantitative analysis, especially in large scale. Considering the immense price reduction and improved mechanical robustness compared to the standard FT-IR instrument, the NLIR instrument is especially attractive for industrial applications. . Note that, the instrument in use was a prototype and a final commercial version will provide an improved spectral quality due to standardization (and optimization) in the optical core.

2.3.6 On-line MIR-spectroscopy

Ultimately aiming at on-line monitoring of the components under study, the NLIR instrument was equipped with an ATR crystal designed as a flow-through cell to facilitate on-line measurements. A photo of the set-up is presented in the supplementary material, Figure S2.5. The fermentation broth was recirculated via the flow-through cell, facilitating automated on-line measurements every 10 minutes over 24 h. Additionally, every 6th sample was complemented with HPLC measurement of the components glucose, ethanol, glycerol and acetic acid. As phosphate and ammonium are not particularly relevant to the fermentation process, their determination was neglected in the on-line approach. 11 manual samples were taken, 10 samples over the first 10 h of the fermentation and a last one after 24 h. The spectral evolution of the flow-through data set is shown in [Figure 2.17 A](#), complemented by a PCA score plot in [Figure 2.17 B](#). The PCA was performed in order to compare

the data sets 1 - 6, measured by means of the off-line ATR, and the data acquired via the flow-through ATR.

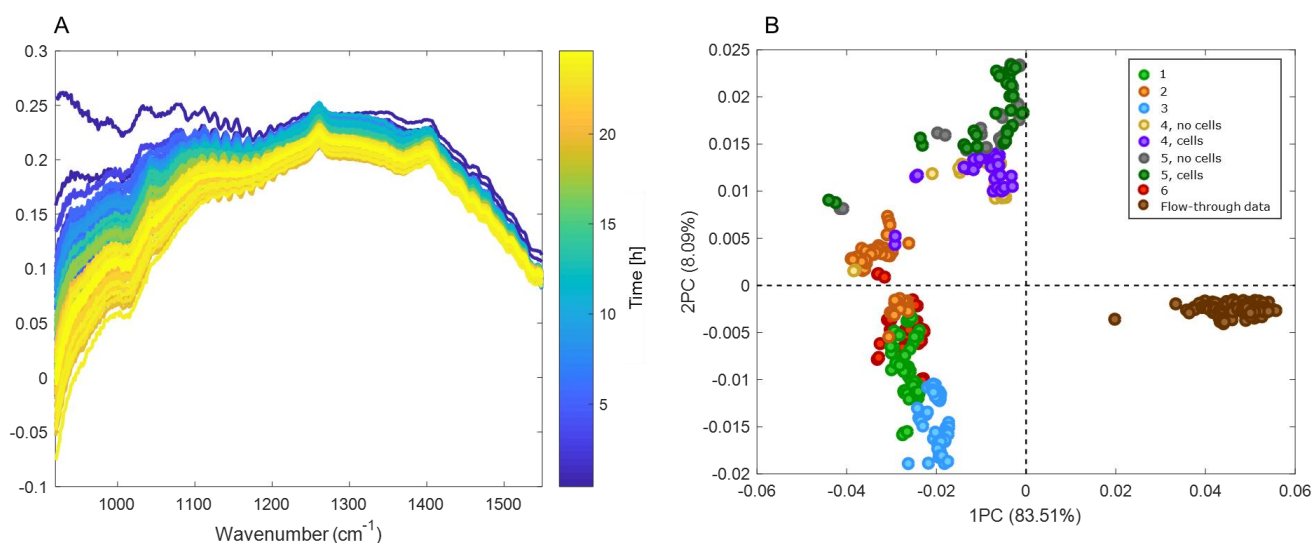




Figure 2.17: A) Spectral evolution of the flow-through data. B) PCA on data sets 1-6 (off-line ATR crystal) and the flow-through ATR crystal.

Figure 2.17 A reveals that the spectral trend of the flow through data is coherent with the spectral trend found in the offline samples (Figure 2.8 A). It furthermore seems that the previous assumption, that the spectral trend is reversed upon a certain point of the fermentation, is confirmed by this set-up, enabling continuous spectra collected between 10 and 24 h. Comparing Figure 2.17 A and Figure 2.9 (left-hand side), two main deviations between the off-line ATR data and the flow-through ATR data become obvious. Firstly, the baseline of the spectra is just the opposite shape, in principle shaping a 'valley' (off-line ATR data) and a 'hill' (flow-through ATR data). This difference resulting from the shape of the baseline separates the flow-through ATR data clearly from the off-line ATR-data along the first principal component, as shown in Figure 2.17 B, explaining more than 83 % of the spectral variation. This difference might be caused by the different characteristics of the ATR crystals used, as summarized in Table 2.7, and /or the application of flow during the on-line measurement.

Table 2.7: Comparison of the flow-through ATR and the off-line ATR

Point of comparison	Off-line ATR	Flow-through ATR
Material	ZnSe crystal, 45°C	
Surface shape		
Surface Area	~ 29 cm ²	~ 14 cm ²
Bounces	12	10

Although composed of the same material, the two ATR crystals differ in shape, surface area and the number of bounces of the IR-light. Thus, the spectral quality is quite different between the off-line data sets 1-6 and the flow-through data set.

Secondly, a peak at around 1240 cm⁻¹ clearly occurs in the flow-through data set (Figure 2.17 A), as it does in the spectra collected off-line on cell-containing samples in Figure 2.9 (right-hand side). A peak at this wavenumber can only be adumbrated in the off-line data of cell-free samples (Figure 2.9, left-hand side). Note that, the fermentation broth was circulated and thus contained cells. The peak at around 1240 cm⁻¹ is assumed to be the spectral fingerprint of the growing yeast cells [46] and peptides from the medium (amide III absorption) [44].

The differences between the spectral off-line and on-line data made the successful application of a model built on off-line data unlikely. In order to confirm this assumption, model 3 was applied to the flow-through spectra that were complemented with HPLC analysis (Figure S2.4 in the supplementary material). With respect to the fermentation course, the flow-through data integrated nicely in the calibration range spanned by model 3, so that deficient predictions could not be linked to an unbalance regarding the calibration and validation space. The fermentation profiles of glucose, glycerol, acetic acid and ethanol, for data set 1 and 6, collected with the off-line ATR (calibration) and the flow-through data (validation) can be found in the supplementary material, Figure S2.3. As expected, and presented in Figure S2.4 in the supplementary material, model 3 applied to the flow-through data seemed to be 'blind'. The RMSEP for the components glucose, glycerol, acetic acid and ethanol lay between 61 % (glucose) and 73 % (acetic acid) and the actual component profile was not captured. The predictions oscillated around a concentration of zero, for all four components.

However, in order to investigate the potential of the on-line set-up, a new model was built (in the following called 'the flow-through model') on the spectra that were supplemented with HPLC data. It was subsequently applied to the spectra collected in between the reference points. The flow-through model and its performance together with parameters relevant to the calibration (number of latent variables, RMSE, bias and R² for calibration (Cal) and cross-validation (CV)) are shown in Figure 2.18 -2.19. Additionally, parameters relevant to calibration are summarized in Table 2.8. The flow-through model was built on data subject to the first derivative as preprocessing technique.

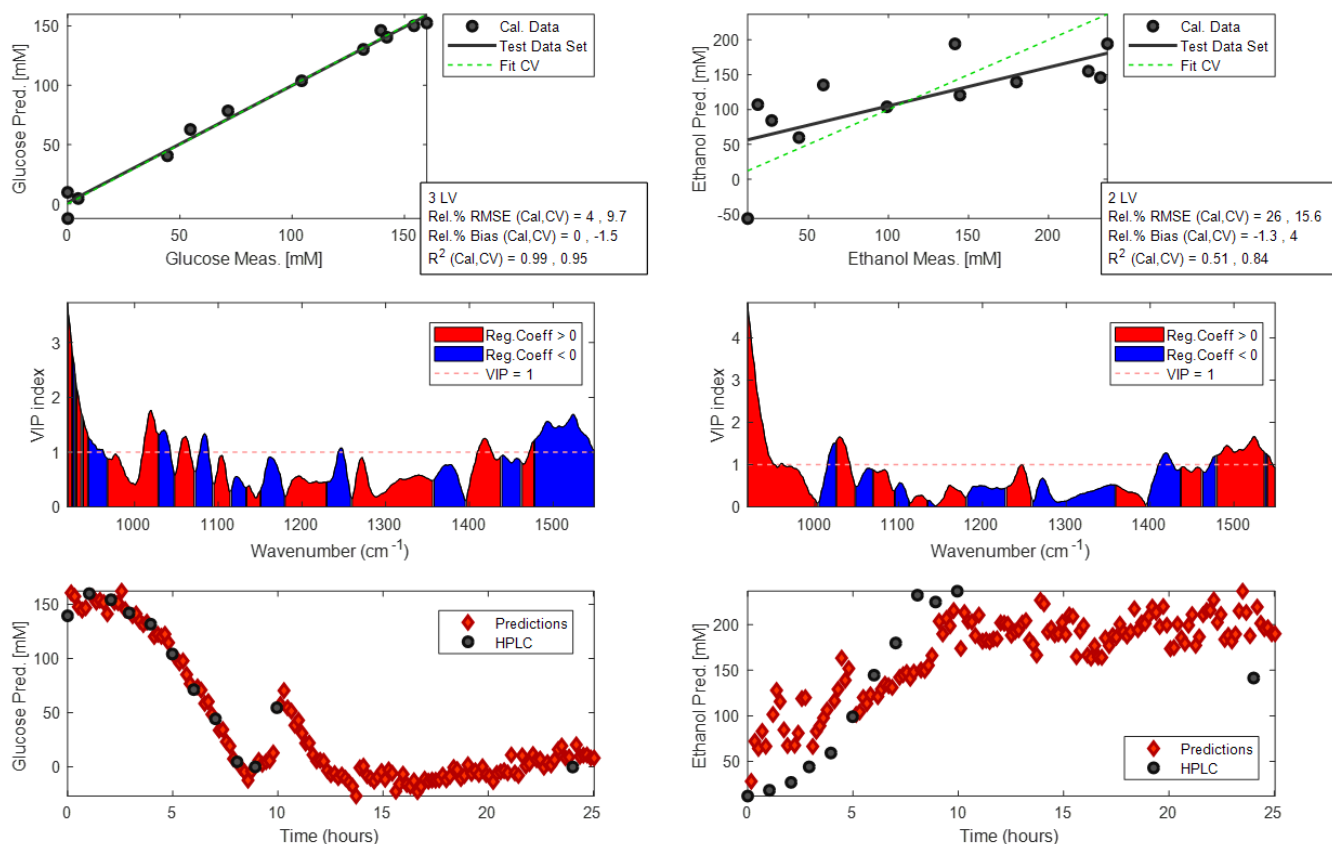


Figure 2.18: Flow-through PLS calibration model and model performance for glucose (left) and ethanol (right). For both components from top to bottom:

PLS calibration model: Additional model parameters relevant to the calibration set as outlined in the box: number of latent variables (LV), relative RMSE, bias and R^2 of calibration (Cal) and cross validation (CV)

VIP index and indication of positive (red) or negative (blue) regression coefficient as a function of wavenumber;

Visual impression of model performance: prediction (red diamonds) vs. HPLC data (black dots) in mM as a function of fermentation time in h. Note that, samples were collected hourly over the first 10 h, and a last sample was collected after 24 h.

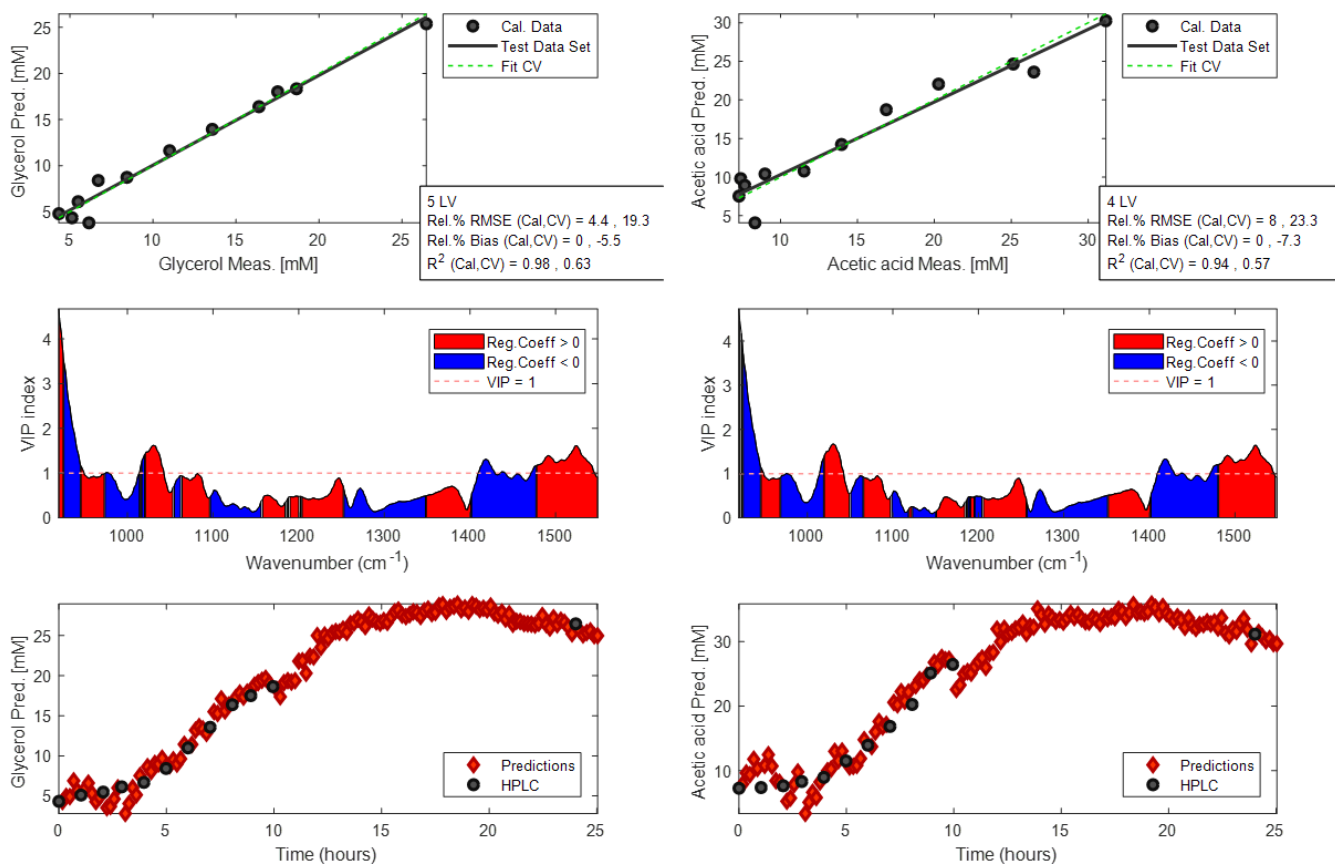


Figure 2.19: Flow-through PLS calibration model and model performance for glycerol (left) and acetic acid (right). For both components from top to bottom:

PLS calibration model: Additional model parameters relevant to the calibration set as outlined in the box: number of latent variables (LV), relative RMSE, bias and R^2 of calibration (Cal) and cross validation (CV)

VIP index and indication of positive (red) or negative (blue) regression coefficient as a function of wavenumber;

Visual impression of model performance: prediction (red diamonds) vs. HPLC data (black dots) in mM as a function of fermentation time in h. Note that, samples were collected hourly over the first 10 h, and a last sample was collected after 24 h.

Table 2.8: Summary of the flow-through-model calibration parameters presented as part of the Figures 2.12-2.14.

Parameter	Glucose	Glycerol	Acetic Acid	Ethanol
Calibration				
12 spectra collected on-line complemented with HPLC data				
Nr. of LV	3	5	4	2
R²(Cal / CV)	0.99 / 0.95	0.98 / 0.63	0.94 / 0.57	0.51 / 0.84
RMSE [%] (Cal / CV)	4 / 9.7	4.4 / 19.3	8 / 23.3	26 / 15.6
Bias [%] (Calibration)	0 / -1.5	0 / -5.5	0 / -7.3	-1.3 / 4

As summarized in Table 2.8 and presented in Figure 2.18 and 2.19, the calibration of the flow-through model resulted in R^2 values of 0.94, 0.98 and 0.99, respectively, for acetic acid, glycerol and glucose. These values are in line with a small bias (max. 7.3 % for acetic acid) and a small RMSE (max. 8 % for acetic acid). Note that, the components glycerol and acetic acid were predicted accurately in this data set, as process correlations were not decoupled. A model like this relying on process correlations may still be relevant and useful in case a stable and reproducible process is in place. The worst calibration was obtained for ethanol yielding an R^2 of 0.51, complemented with a RMSE(Cal) of 26 %, however, a small bias of 4 %. Cross-validation yielded certainly lower R^2 values than the calibration, in particular for glycerol and acetic acid. In both cases, the R^2 (CV) is around 0.6 resulting in a RMSECV of around 20 %. The models for glucose and ethanol yielded better values for cross-validation (R^2 (CV)), 0.95 and 0.84, respectively, reflecting a RMSECV of 10 % and 15 %. All models were built on a significantly lower number of latent variables compared to the models built on the off-line ATR data (compare to Table 2.5), maximal 5 for glycerol, and then 4, 3, and 2 for modelling acetic acid, glucose and ethanol. Models 1-8 built on the off-line ATR data were built upon 3 – 10 latent variables (exemplary demonstrated in Table 2.5). This is in agreement with the prior investigations of the flow-through data performed by PCA analysis (Figure 2.17 B) and reveals once again that the spectral quality significantly differs between the two ATR elements used. The number of latent variables generally suggests a greater model robustness for the flow-through data, compared to the off-line data.

As advocated by the calibration parameters, the predictions perform really well, in particular for the components glucose, glycerol and acetic acid, while the prediction of ethanol seems a bit more challenged (Figure 2.18 and 2.19). In case of the three former analytes, the number of spectra collected (every 10 min over 24 h) yielded a continuous trace resulting in a reliable component profile which is in very good agreement with the reference data points. Besides, the spiking of glucose after 9 h, simultaneously diluting the fermentation broth, is captured accurately by the three models. It causes in a prediction peak and a drop, respectively, in the glucose and glycerol / acetic acid profiles. It is amazing how the fermentation profile between 10 and 24 h could be recovered, without a single reference measurement in between. The resulting curve looks reasonable and is in agreement with the expected metabolism. The results are very promising regarding future predictions.

It must be emphasized that the flow-through ATR as such and connected to the NLIR instrument was used for the very first time within this work. The approach demonstrated that the flow-through ATR connected to the fermenter enabled fast and efficient spectral sampling and the first modelling attempt (Figure 2.18 and 2.19) was very promising. However, a tailored calibration model must be built on this particular set-up. Synthetic samples must be included in order to decouple the correlations between the process parameters, aiming at robust models. However, in case a strong (reproducible) process is in place, the natural process correlations strongly improved the modelling of glycerol and acetic acid, and decoupling the process dynamics is not necessary here. In fact, such a model would be more accurate, however less robust. Hence, more attention and care must be dedicated to model maintenance. However, the possibility of on-line spectra acquisition quite simplifies this task, as may be three off-line samples providing reference analysis might be enough to recalibrate the model, or check, if the model still works. Further investigations regarding model maintenance and related practical issues such as the influence of the flow rate itself, the direction of flow, the creation of a biofilm on top of the ATR surface over time and a proper cleaning procedure are to be considered.

2.3.6.1 Influence of the flow-rate on on-line spectra collection

The influence of the flow rate on the spectrum was studied in a first attempt with water. A flow of 0.5 ml/min to 10 ml/min was investigated, changing the flow rate in steps of 1 ml/min, starting from a low flow rate of 0.5 ml/min, then 1ml/min, 2 ml/min and so forth. The flow direction was in-line with the beam of the IR light source. Every flowrate was studied over 3 min, collecting a spectrum every 20 seconds. After the change of the flow rate, 1 minute waiting time was taken into account before starting the spectra collection. The results are presented in Figure 2.20.

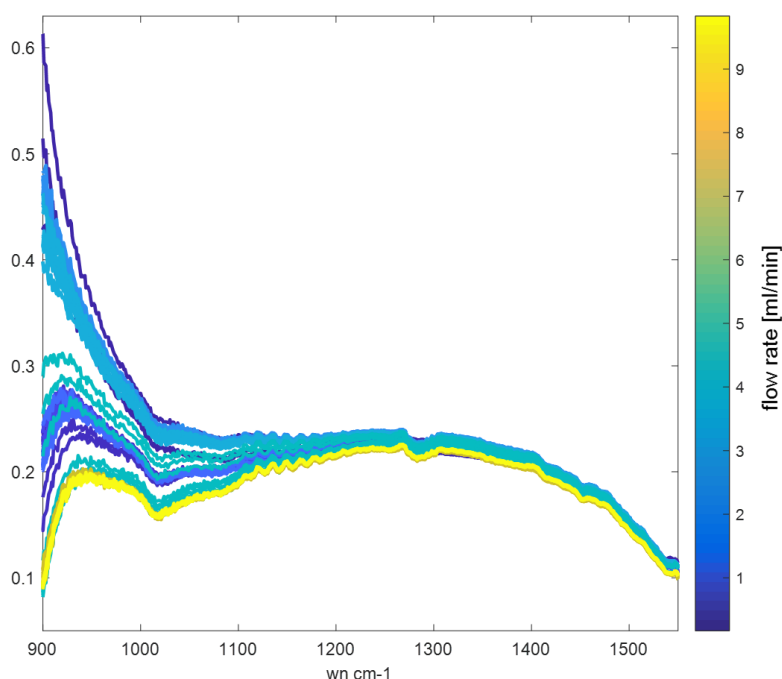


Figure 2.20: Influence of flow rate on the MIR spectra acquired with water. A flow rate range of 0.5 to 10 ml/min was investigated in steps of 1 ml/min starting with 0.5 ml/min, then 1ml/min, 2 ml/min and so forth as indicated by the color gradient in the figure. Every flowrate was studied over 3 min, collecting a spectrum every 20 seconds. After the change of the flow rate, 1 minute waiting time was taken into account before starting the collection of spectra.

It was found that the spectral baseline was quite dependent on the flow rate, in particular in the range of 900 – 1100 cm^{-1} (referred to as 'lower region'). The region above 1100 cm^{-1} was affected to a lower extent, yet the effect was visible. Furthermore, the flow rate seems to influence the reproducibility, again especially in the lower region. While 1 ml/min resulted in broad spectral deviation in the lower region, a flow rate of 10 ml/min resulted in entirely and over the full spectral range overlapping spectra. A flow rate of 5 ml/min, as applied during the continuous on-line set-up, clearly resulted in spectral deviation between the different measurements of the same flow rate, especially in the lower spectral region. It is important to notice that ethanol as a pure component yields absorption bands in especially that region, and exhibits the overall lowest molecular absorption across the components causing absorption bands within this lower region (glucose, glycerol, ethanol and phosphate, [Figure 2.7](#) and [2.8](#)). It explains why the calibration of ethanol resulted in the overall lowest R^2 value: Under this flow rate, the reproducibility is poor especially in the lower region containing the spectral fingerprint of ethanol and the model is furthermore challenged by low molecular absorption of this component. A final observation is important to mention. The baseline generally is quite noisy and the peak at 1240 cm^{-1} previously discussed to be caused by cells, is clearly visible in all the water spectra collected. It must be concluded that the previous assumption must be handled with caution. Yet, it is not entirely proven wrong as the influence of the flow rate was tested right after the fermentation and although the cell was thoroughly flushed with water and 70 % ethanol, it is not clear if traces of the fermentation remained on the ATR surface. This also points out that a proper cleaning procedure must be developed including a standard spectrum of a 'clean ATR'.

2.3.6.2 Pure component spectra acquired on-line

In order to evaluate the RMSEP obtained for the models built upon spectral data acquired under flow-conditions (summarized in [Table 2.8](#)), the pure component spectra of the different analytes were collected under flow-conditions as well, at a concentration of 0.1M and 1 M. The individual and overlaid results are presented in [Figure 2.21 A](#) and [2.21 B](#), respectively.

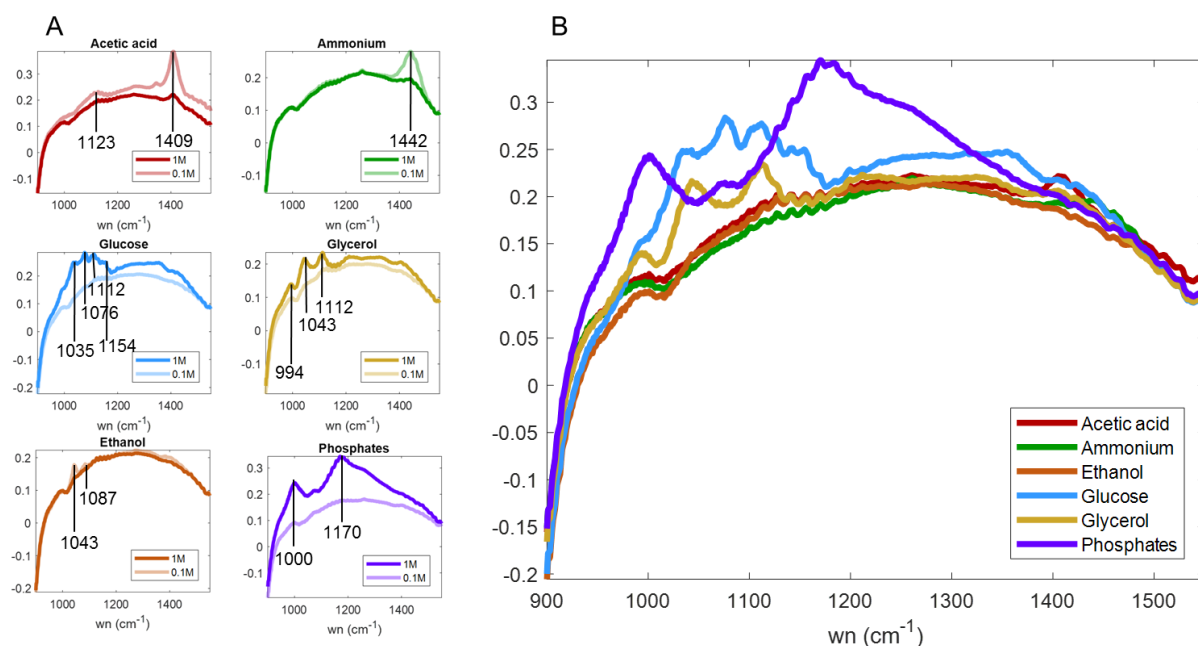


Figure 21: IR spectra of pure components in Milli Q water, collected with the NLIR instrument equipped with the flow-through ATR. The flow rate was 0.5 ml/ min. A) Each spectrum shows the absorbance as a function of wave number (wn) in cm^{-1} of 2 different analyte concentration as indicated in the legends (0.1 and 1 M). B) Pure component spectra of the 1 M solution in an overlaid fashion.

The absorption bands identified in the NLIR spectra acquired via the flow-through ATR are summarized in Table 2.9 and compared to the absorption bands identified in the FT-IR spectra of the pure components.

Table 2.9: Absorption peaks identified in cm^{-1} ; FT-IR vs. Flow-Through NLIR; The peak assignment refers to Figure 2.7 (FT-IR) and Figure 2.21 (NLIR, flow-through ATR).

Glucose	Glycerol	Acetic Acid	Ethanol	Ammonium	Phosphate
1034 / 1035	- / 994	- / 1123			
1080 / 1076			1045 / 1043		1076 / 1000
1105 / 1112	1041 / 1043	1277 / -	1078 / 1087	1456 / 1442	1157 / 1170
1151 / 1154	1111 / 1112	1414 / 1409			

Both studies are overall coherent but small differences are present, in particular for the components glycerol and acetic acid. An extra peak and a clear shift in peak is observed, respectively. It is obvious that the spectra acquired under flow with the NLIR instrument (Figure 2.21) are more noisy and less sensitive than the FT-IR spectra (Figure 2.7 and 2.8 A), and off-line ATR spectra (Figure 2.9). It is striking that a concentration of 1 M yielded an absorption not higher than 0.3 with the NLIR instrument equipped with the flow-through ATR, whereas a concentration of 0.2 M yielded an absorption of about 0.7 with the FT-IR instrument. Especially ethanol yielded marginal absorbance in the flow-through set-up. The low sensitivity of the flow-through ATR is also revealed when comparing Figure 2.9 (spectral evolution of off-line fermentations samples) and Figure 2.17 A (spectral evolution of flow-through fermentation samples). The off-line set-up resulted in an average absorption of 0.55, while the flow-through set-up resulted in an absorption of on average 0.2. However, these differences must be attributed to the different ATR elements used.

It can be concluded that the flow-through set-up was considerably less sensitive compared to the FT-IR and off-line-ATR-NLIR set-up. The somewhat worse model obtained for ethanol under flow-conditions can be explained by the very low IR-activity of this component, explaining why the ethanol signal was somehow suppressed in the flow-through set-up. The low(er) signal-to-noise ratio in the relevant region in the flow-through set-up was found to challenge the modeling of this component.

2.3.6.3 Workflow for PLS-model development under flow-conditions

It was demonstrated that models based on continuous spectra collection have a great potential to be transferred into an on-line concentration profile of the respective component. With the knowledge gained until now and the necessity for a calibration model applicable to this set-up, a calibration procedure for the application under flow is suggested as explained below. For this, the flow-through set-up and fermentation process under study 'as is', are considered. A similar workflow is presented in [16], while this one complements the cited work with respect to the identification of indirect relationships, and the consideration of flow-conditions.

1) Choice of the flowrate

A flow rate of 10 ml/min seems to be suitable regarding spectral reproducibility and appears to be applicable for operation. However, this should be investigated and proven with the actual fermentation matrix (YP medium). The calibration and operational measurements are to be performed under the same flow rate. The transfer of a model obtained at a specific flow-rate to another flow rate must be investigated. It might be considered that an off-line ATR equal to the ATR used in the on-line set-up might be beneficial. In such a case, the effect of flow can be studied as a matter of direct comparison. Eventually, spectra collected off- and on-line could be combined into a 'master' calibration set. In such a way, process and synthetic samples could be acquired in a time effective and practical way. However, this must be investigated with a respective off-line and on-line ATR.

2) Acquisition of the pure component spectra for the chosen flow rate

Ideally, an off-line ATR equal to the flow-through ATR is available. Thus, pure component spectra might also be acquired off-line as discussed in 1). The pure component spectra should be acquired at least at two, better three different concentrations, e.g. 20mM, 100 mM and 200 mM, in water and the fermentation matrix. If the actual concentration range of the relevant process parameters is known, pure component spectra could be acquired at the minimum and maximum level present during operation. However, with regards to a consistent library of pure component spectra, a general min. and max. concentration should be acquired as well. By means of the pure component spectra, knowledge about the IR activity and absorption bands of the pure components is gained. Furthermore, challenges arising from instrumentation, overlapping spectral features and the effect of the fermentation matrix can also be evaluated. Reagents as acid and base for pH control should also be taken into account when studying the pure components. This knowledge will be valuable when evaluating the models built. Finally, a library of pure component spectra can be implemented, and relevant concentration levels can later on be included 'right away' into the calibration data.

3) Model calibration

The quality and design of the calibration samples is of crucial importance in order to build a strong calibration model. Especially in biotech based research, multivariate calibration is a topic poorly addressed in literature and general guidelines are missing. Important key points regarding the topic of multivariate calibration are outlined in detail by Andrey Bogomolov [47]. Thus, his publication is to be mentioned here. It suggests and provides a Matlab script for a so called diagonal design, not limited by the number of components to be calibrated and considers all the important aspects connected to this subject. The suggested design might be interesting to consider for future calibration designs.

3.1) Collection of fermentation samples

In order to obtain a robust calibration model, relevant process conditions and eventual disturbances should be considered and included in the calibration data. Ideally, each fermentation is run as a duplicate providing the possibility for a representative external validation set. It is important to consider that the calibration model built will only account accurately for variation included in the calibration data. A number of samples relevant to the process dynamics are to be chosen. Within this work, 15 samples generally assembled the profiles of the batch phase well and moreover accounted for the overall concentration change of each component during the process. Spectra and reference analysis of the fermentation samples should be performed in duplicates. The quality of the reference analysis directly affects the calibration model and should be consistent, reproducible and accurate. Note that, the accuracy of the calibration model can never be better than the accuracy of the reference analysis, the smallest concentration change calibrated and the signal to noise ratio (SNR) of the spectroscopic system in use. Calibration can be evaluated based on R^2 , bias and RMSE of calibration and cross-validation. The performance of cross-validation might be overoptimistic with respects to future samples but can be used as an indication of challenged predictions (see chapter 2.3.4.3 Detailed presentation of Model 3a and [Table 2.5](#)).

3.2) Spiking of fermentation samples

Spiking of fermentation samples can be used in order to decouple the correlations between the different analytes. Besides, spiked samples are a good measure of model robustness, in other words, if spectral variation is directly connected to the analyte of interest. If the model does not accurately respond to the spiking, it is likely built on indirect relationships. Also here, it is important that the spiking of process –relevant parameter concentrations is done within the relevant process frame. Spiked samples exceeding the calibration space cannot be expected to be extrapolated accurately by the model and might lead to misleading conclusions about the model performance.

3.3) Decoupling the process correlations

Generally, the substrates and products involved in a fermentation process are inherently trapped in a biological cave of covariance. Hence, it is important to understand and to decouple underlying correlations in order to yield a targeted PLS calibration model. Note that, in case of models built upon indirect relationships, decoupling of correlations decreases the model performance. However, the correlations between the different components might be studied and decoupled as outlined within this thesis, section 2.2.1.2 Design of synthetic samples. It will be beneficial to equally space the concentration values within the minimum and maximum value considered. The minimum concentration step within the synthetic samples is considered to determine the sensitivity of the

calibration model. The number of synthetic samples should equal or extend the number of fermentation samples needed to span a reasonable component profile. A higher number might be beneficial. Besides, it was demonstrated that components produced by the microorganism but not measured during the fermentation process, (within this study e.g. lactic acid) might be included in the synthetic samples. They account for valuable and relevant process variability and their addition into the calibration set was shown to improve the PLS model in [36]. It is interesting to mention that relevant but unmeasured metabolites might be identified and tracked by means of a residual plot as demonstrated in [38].

3.4) Data exploration by PCA analysis

A PCA of the spectral fermentation data can give valuable information about the batch to batch variability as well as the chemical rank of the fermentation spectra. The latter is a good indicator of how many components can be modeled based on independent latent variables.

4) Model validation

In order to get a realistic idea of the model performance, validation should always be performed on an external validation set, that is to say samples that are not part of the calibration set. It might be beneficial to include batches next to the duplicates referring to the calibration data, and evaluate the model performance overall and per data set. Model performance can be estimated by the relative root-mean-square-error (RMSEP) and bias of the predicted data. It must be emphasized that the RMSEP should not be used as the only reference when evaluating PLS model performance. Predictions based on indirect relationships are not directly captured by the PLS model and must be investigated separately.

4.1) Identification of indirect relationships

If a strong calibration procedure is in place, the calibration space is properly spanned for each component and the process correlations are decoupled by synthetic and/ or pure component samples. Then, the overlapping spectral features must be taken into account. This information is not captured by the PLS model and must be evaluated separately. As a conclusion resulting from this work, overlapping spectral features do compromise in particular the prediction of minor components with low molecular absorption. They are masked by the strong absorption of major components within the same spectral region. However, if components with low molecular absorption are present in reasonable amounts even in regions of overlapping absorption bands, they could be modelled successfully.

Cross-correlations are indicated by high R^2 values and high VIP scores related to the same spectral region, as presented in [Figure 2.15](#). However, this basically complements the initial study of the pure component spectra and the interpretation of these parameters might not lead to a clear result. Indirect relations can be quali- and quantitatively identified by means of the correlation with a certain prediction as performed in this work ([Figure 2.16](#)), as well as by the selectivity ratio [48] and explained variance [24], [23] for each component, described elsewhere.

2.4 Conclusions and Future Perspective

As demonstrated in this work, IR-spectroscopy combined with PLS-modelling is a highly powerful tool for rapid multi-component quantification in the complex fermentation matrix. However, the calibration process is complex and intense and might only pay off on the long-term view, when providing consistent process conditions. This applies to industrial production where bioprocesses must be operated under optimized and defined conditions. In this environment, IR-spectroscopy combined with PLS modelling is expected to be a strong tool, ensuring consistent processing and quality of biotechnological products. The novel technology employed within this study, called NLIR, scored with increased instrumental robustness and flexibility compared to a standard FT-IR lab instrument. The spectral analysis in the NLIR instrument is tailored towards a narrow region of so-called fingerprints, occurring between 800 cm^{-1} – 1800 cm^{-1} . The particular application range is designed for the region 900 cm^{-1} – 1550 cm^{-1} , avoiding the strong water absorption at 1630 cm^{-1} . Yet, the applied region is rich in spectral information related to the functional groups comprising most bio-molecules. The spectral downgrading was not found to be significant compared with the standard FT-IR instrument employed. In both cases, models for glucose, ethanol, ammonium and phosphate yielded a RMSEP of 8.6 % and 10 % on average (external validation), the latter accounting for the NLIR instrument. With its practical features involving an immense price reduction compared to the standard FT-IR instrument, the NLIR instrument is especially attractive for the application in large scale production. Models built on off-line data yielded errors of less than 15 %, the relative RMSEP values for glucose, ethanol, ammonium and phosphate gave 3.8 %, 13.2 %, 9.5 % and 7.9 %, in the best case (Table 2.4). Within this work, glucose, ethanol, and to a lesser extent ammonium and phosphate were modeled based on distinct spectral features. The calibration of glycerol and acetic acid was challenged by low concentration levels within the process, relatively low IR activity compared to the other components, and cross-correlation of spectral absorption bands. Thus, PLS-calibration models for glycerol and acetic acid were shown to rely on indirect prediction, trapped in a cave of covariance with ammonium and phosphate, or respectively, regions linked to the spectral absorbance of ammonium and phosphate (Figure 2.16). Consequently, RMSEP values for both models yielded 30 % - 200 %, depending on how well calibration and validation space were matching. However, both components can potentially be modelled reliably in another process set-up. Generally, it was discussed that the calibration procedure could be optimized by a calibration design including equal spacing of the synthetic samples, an improved calibration range better representing the process variation, and eventually, by performing a variable selection. This study pointed out the importance of multivariate calibration design and the evaluation of obtained PLS-models beyond the common calibration and validation parameters as R^2 , bias and RMSE.

Shown to have the potential to overcome the lack of monitoring strategies in the fermentation area, the application of IR-spectroscopy and PLS-modelling is challenged by a reasonable amount of expertise needed in different fields accompanied with a lack of general guidelines, methodologies and software. By now, novel technology has addressed issues related to instrumentation. The instrument in use could be modified as a flow-through unit, enabling the frequent, automated collection of spectra, and a first modelling approach of this data was very promising. However, a new calibration procedure must be taken into account, since the ATR-element used in the on-line setup differs from the ATR used for off-line model development. Besides, the application of flow during the acquisition of spectra was found to alter the spectral baseline.

The development of a strong and robust PLS-model is data intense and hence, time and labor expensive. Besides, models obtained for one process might not be transferred to another process due to a different process environment. In other words, modelling the same component in different processes might indeed yield fairly different results. Nevertheless, investing time and money in the beginning will pay off on the long -term. The ongoing development of novel technology offers the flexibility needed to comply with so far neglected needs. In order to be outstanding, the novel instrumentation can and should be complemented with a proper software including all the tools needed to turn the measurement method into an applicable soft sensor. A library of pure component spectra in water and relevant media can be started, so that the study of absorption bands and overlapping spectral features becomes faster and faster in different set-ups. Furthermore, the software can either contain chemometric and calibration design tools or should be able to communicate with common programming languages (as Matlab, Python and R), as well as common data storage files (excel, text). When containing the respective tools, data could be directly plotted and investigated. Studying and decoupling the process dynamics aiming at the design of synthetic samples according to multivariate calibration standards, and PLS modelling would be facilitated 'at the same place'. For instance, the software could comprise 3 modules: one for online spectra collection, one for general and chemometric off-line analysis and one for on-line spectra collection and model application. An interdisciplinary interface that facilitates a platform complementing the different areas, spectral investigation, displaying and decoupling of process correlations, multivariate calibration design and PLS (chemometric)-modeling, might be the key to successful industrial application.

2.5 Supplementary Material

2.5.1 Tables

Table S2.1: Composition of synthetic samples in mM.

sample	Glucose [mM]	Glycerol [mM]	Acetic acid [mM]	Ethanol [mM]	Ammonium [mM]	Phosphate [mM]
1	180	35	5	30	70	45
2	0	5	5	10	95	30
3	65	30	15	65	55	80
4	90	30	25	20	30	70
5	50	45	10	140	50	5
6	15	20	15	95	80	100
7	105	10	20	130	15	50
8	40	0	40	75	35	35
9	130	20	35	120	85	15
10	155	45	0	85	100	85
11	140	40	35	0	5	20
12	165	50	30	150	0	0
13	115	5	30	105	65	55
14	75	15	10	45	20	95
15	25	25	25	55	45	65

Table S2.2: Spiking scheme for the preparation of the synthetic samples based on a 2 M stock solution at pH 6 for each component. The given volumes were added to a matrix of YP medium at pH 6 to achieve a total sample volume of 15 ml.

sample	Glucose [ml]	Glycerol [ml]	Acetic acid [ml]	Ethanol [ml]	Ammonium [ml]	Phosphate [ml]
1	1.350	0.263	0.038	0.225	0.525	0.338
2	0	0.038	0.038	0.075	0.713	0.225
3	0.488	0.225	0.113	0.488	0.413	0.600
4	0.675	0.225	0.188	0.150	0.225	0.525
5	0.375	0.338	0.075	1.050	0.375	0.038
6	0.113	0.150	0.113	0.713	0.600	0.750
7	0.788	0.075	0.150	0.975	0.113	0.375
8	0.300	0	0.300	0.563	0.263	0.263
9	0.975	0.150	0.263	0.900	0.638	0.113
10	1.163	0.338	0	0.638	0.750	0.638
11	1.050	0.300	0.263	0.000	0.038	0.150
12	1.238	0.375	0.225	1.125	0	0
13	0.863	0.038	0.225	0.788	0.488	0.413
14	0.563	0.113	0.075	0.338	0.150	0.713
15	0.188	0.188	0.188	0.413	0.338	0.488

Table S2.3: Spiking scheme in ml for the preparation of the spiked samples (data set 5, Table 2.1, main text). Spiking was performed with a 2 M stock solutions at pH 6 for each component. The given volumes were added to each fermentation sample to achieve a final sample volume of 10 ml. The number of each sample was tied to the hour of sampling.

sample	Glucose [ml]	Glycerol [ml]	Acetic acid [ml]	Ethanol [ml]	Ammonium [ml]	Phosphate [ml]
0	0.000	0.025	0.175	0.725	0.025	0.175
1	0.000	0.000	0.050	0.000	0.325	0.325
2	0.000	0.075	0.050	0.475	0.150	0.150
3	0.000	0.000	0.000	0.550	0.100	0.025
4	0.450	0.000	0.025	0.000	0.075	0.475
5	0.000	0.075	0.125	0.225	0.175	0.450
6	0.525	0.075	0.050	0.000	0.500	0.225
7	0.650	0.050	0.000	0.050	0.275	0.300
8	0.600	0.000	0.025	0.000	0.400	0.375
9	0.150	0.000	0.075	0.000	0.250	0.025
10	0.325	0.050	0.000	0.000	0.475	0.100
11	0.475	0.100	0.000	0.000	0.000	0.000
12	0.700	0.000	0.125	0.375	0.425	0.125
13	0.425	0.075	0.125	0.725	0.200	0.275
14	0.225	0.125	0.075	0.325	0.350	0.425

Table S2.4: Synthetic samples designed based on the minimum and maximum value found (marked in grey) for each component across the fermentations performed. In this design, equal spacing between the 15 samples was considered.

sample	Glucose [mM]	Glycerol [mM]	Acetic acid [mM]	Ethanol [mM]	Ammonium [mM]	Phosphate [mM]
1	0.00	0.00	0.00	0.00	0.00	0.00
2	12.86	3.57	2.86	10.71	7.14	7.14
3	25.71	7.14	5.71	21.43	14.29	14.29
4	38.57	10.71	8.57	32.14	21.43	21.43
5	51.43	14.29	11.43	42.86	28.57	28.57
6	64.29	17.86	14.29	53.57	35.71	35.71
7	77.14	21.43	17.14	64.29	42.86	42.86
8	90.00	25.00	20.00	75.00	50.00	50.00
9	102.86	28.57	22.86	85.71	57.14	57.14
10	115.71	32.14	25.71	96.43	64.29	64.29
11	128.57	35.71	28.57	107.14	71.43	71.43
12	141.43	39.29	31.43	117.86	78.57	78.57
13	154.29	42.86	34.29	128.57	85.71	85.71
14	167.14	46.43	37.14	139.29	92.86	92.86
15	180.00	50.00	40.00	150.00	100.00	100.00
steps between samples	12.86	3.57	2.86	10.71	7.14	7.14

Table S2.5: Permuted column values based on Table S2.4. The result of the design is presented in Figure S2.2.

sample	Glucose [mM]	Glycerol [mM]	Acetic acid [mM]	Ethanol [mM]	Ammonium [mM]	Phosphate [mM]
1	64.29	17.86	28.57	128.57	35.71	85.71
2	154.29	42.86	31.43	21.43	28.57	100.00
3	12.86	10.71	14.29	75.00	71.43	35.71
4	167.14	25.00	20.00	150.00	64.29	64.29
5	38.57	21.43	37.14	107.14	57.14	42.86
6	115.71	28.57	2.86	139.29	7.14	0.00
7	77.14	14.29	11.43	42.86	0.00	21.43
8	128.57	35.71	0.00	53.57	42.86	57.14
9	51.43	7.14	8.57	85.71	21.43	28.57
10	90.00	46.43	22.86	10.71	92.86	7.14
11	141.43	3.57	25.71	117.86	100.00	92.86
12	0.00	50.00	34.29	64.29	14.29	71.43
13	25.71	39.29	17.14	96.43	78.57	78.57
14	180.00	0.00	5.71	32.14	85.71	14.29
15	102.86	32.14	40.00	0.00	50.00	50.00

Table S2.6: Reference values of data set 4 and 5 (defined in the main text in Table 2.1). The reference data for cell-free and cell-containing samples are the same. Values marked in red are concentration outliers with respect to model 3a (Main text, Table 2.3)

Data set 4	Glucose	Glycerol	Acetic acid	Ethanol	Ammonium	Phosphate
time [h]	[mM]	[mM]	[mM]	[mM]	[mM]	[mM]
0	130.17	3.27	0.00	0.00	0.7	14.3
1	129.74	3.31	0.00	0.00	0.6	14.6
2	128.90	3.36	0.00	0.00	0.7	14.2
3	127.61	3.45	0.00	4.50	0.6	15.1
4	126.02	3.59	0.00	7.47	0.7	13.5
5	122.38	3.82	7.91	16.28	0.6	11.8
6	116.66	4.25	8.32	25.42	0.6	12.8
7	108.12	4.74	8.62	39.22	0.6	12.2
8	94.72	5.55	9.34	61.64	0.6	12.1
9	73.75	6.75	10.53	95.46	0.6	12.2
10	42.59	8.42	12.68	146.74	0.6	11.1
11	4.48	10.41	15.53	211.40	0.6	8.2
12	0.00	10.69	18.67	208.84	0.5	8.3
13	0.73	11.26	20.56	200.92	0.6	7.7
14	0.00	11.30	20.87	193.24	6.2	6.4
26	0.00	10.96	12.26	129.58	13.8	5.1

Data set 5	Glucose	Glycerol	Acetic acid	Ethanol	Ammonium	Phosphate
time [h]	[mM]	[mM]	[mM]	[mM]	[mM]	[mM]
0	115.26	8.01	42.98	154.44	0.8	52.0
1	119.04	3.05	16.97	94.87	50.1	53.8
2	119.46	3.13	17.44	0.00	78.4	69.9
3	119.25	3.25	6.89	116.60	33.5	13.4
4	204.98	3.22	11.96	8.54	27.9	120.9
5	108.57	18.33	40.11	60.22	49.2	121.4
6	208.83	19.02	17.29	21.56	119.7	81.0
7	226.12	14.15	7.00	44.24	68.7	99.3
8	202.24	4.76	13.09	52.80	90.4	116.3
9	99.74	6.32	9.81	105.43	64.3	24.2
10	104.94	17.84	11.24	132.68	107.3	36.5
11	100.29	30.46	14.35	198.76	11.9	2.2
12	143.97	8.70	40.84	251.33	96.7	42.5
13	86.41	24.50	41.94	0.00	50.1	89.6
14	46.96	35.77	32.47	0.00	77.3	87.8
26	96.36	30.76	11.37	0.00	0.8	4.3

2.5.2 Figures

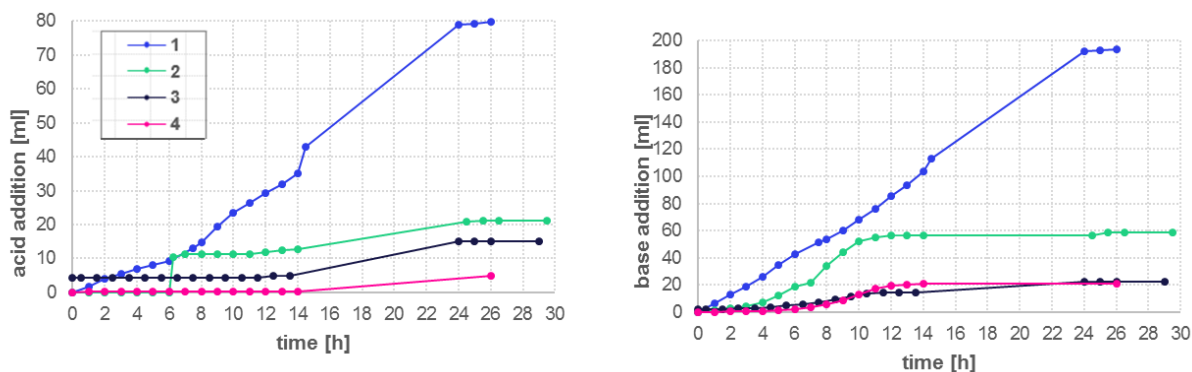


Figure S2.1: Acid and base addition profile for the fermentation 1-4 as outlined in Table 2.1 in the main text. Note that the value of the initial value for acid and base addition might be misleading. The dose monitor was not consistently 'stet to zero'. The value can be considered as zero when the fermentation was started. Acid and base was added by manual pump operation in order to adjust the pH to 6 at the prior to the fermentation start. The pre-addition was unfortunately not always noted. For fermentation 1, the pH controller boundaries were slightly increased compared to the other fermentations. The initial boundaries were found to be too large, resulting in constant overshooting in base addition consequently leading to acid addition. No contamination was observed under the microscope. The settings were changed afterwards and kept constant for all other fermentations.

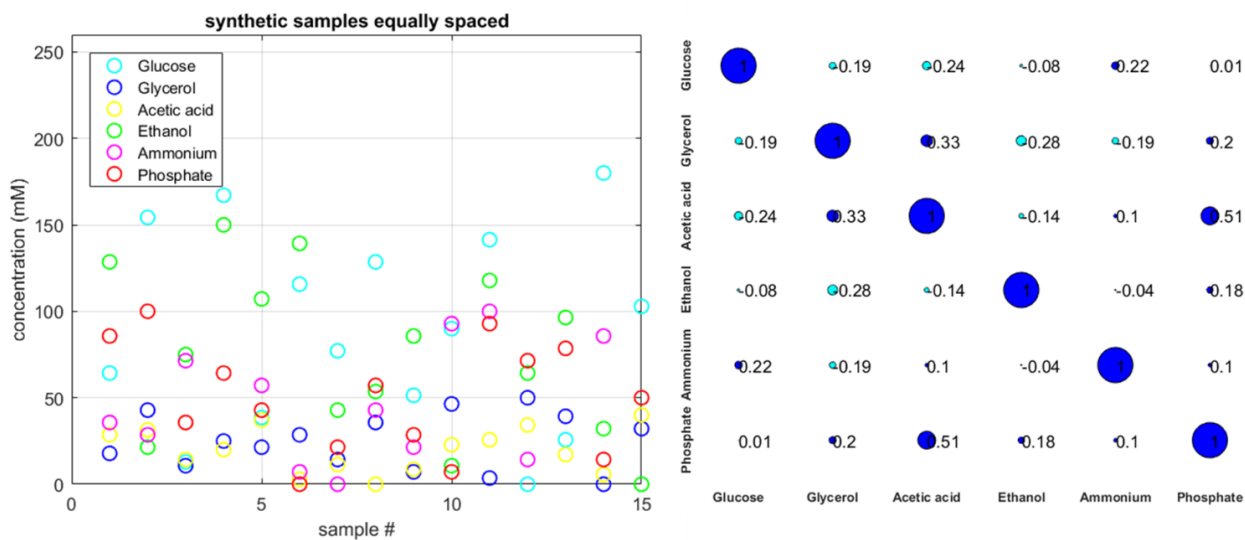


Figure S2.2: Left: Permuted sample composition of the 15 synthetic samples designed based on equal spacing (Table S2.5). Right: Bubble plot of the r values between the different components. A dark blue bubble color indicates a positive correlation, a light blue bubble color indicates a negative correlation between two components. The bubble size indicates the strength of the underlying correlation. The bubbles lying on the diagonal can be seen as a reference for the bubble size of an r value equal to 1 (100 % correlated). The average correlation between the components is 0.19.

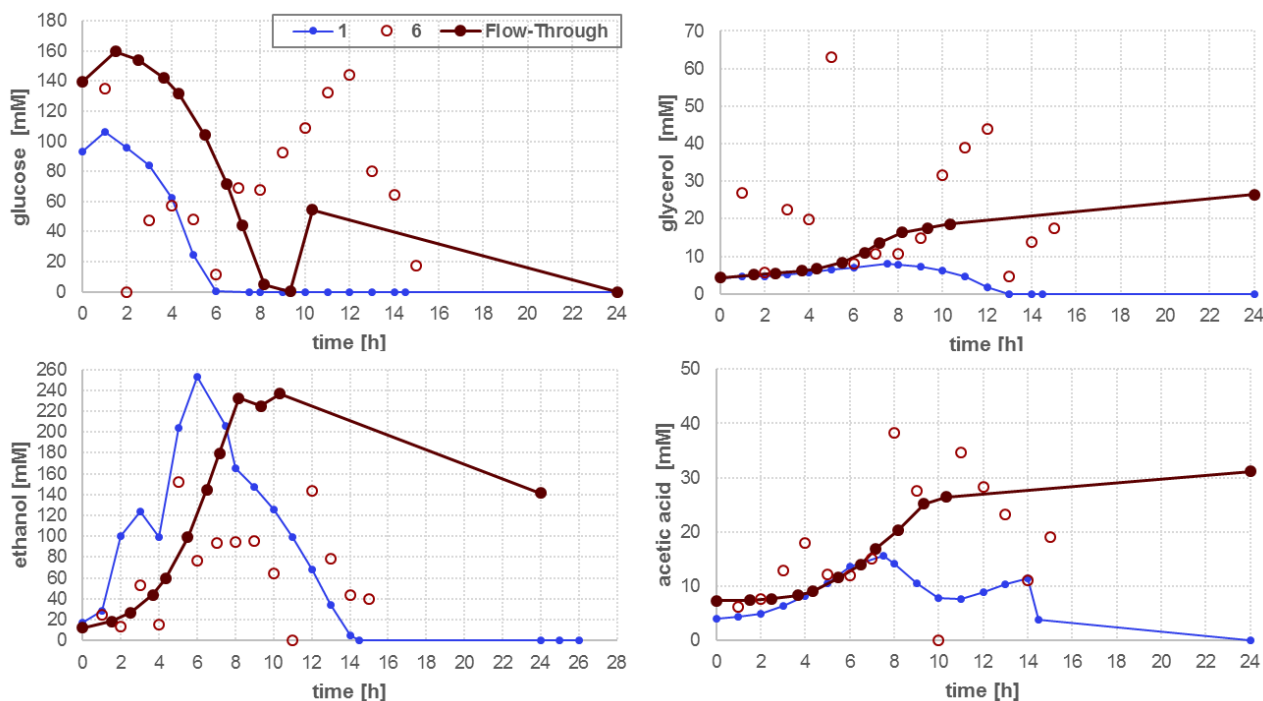


Figure S2.3: Fermentation profiles for the components glucose, glycerol, acetic acid and ethanol for data set 1 and 6 as described in the main text, Table 2.1 and the flow-through data. The legend shown for the analyte glucose holds for all the other components as well.

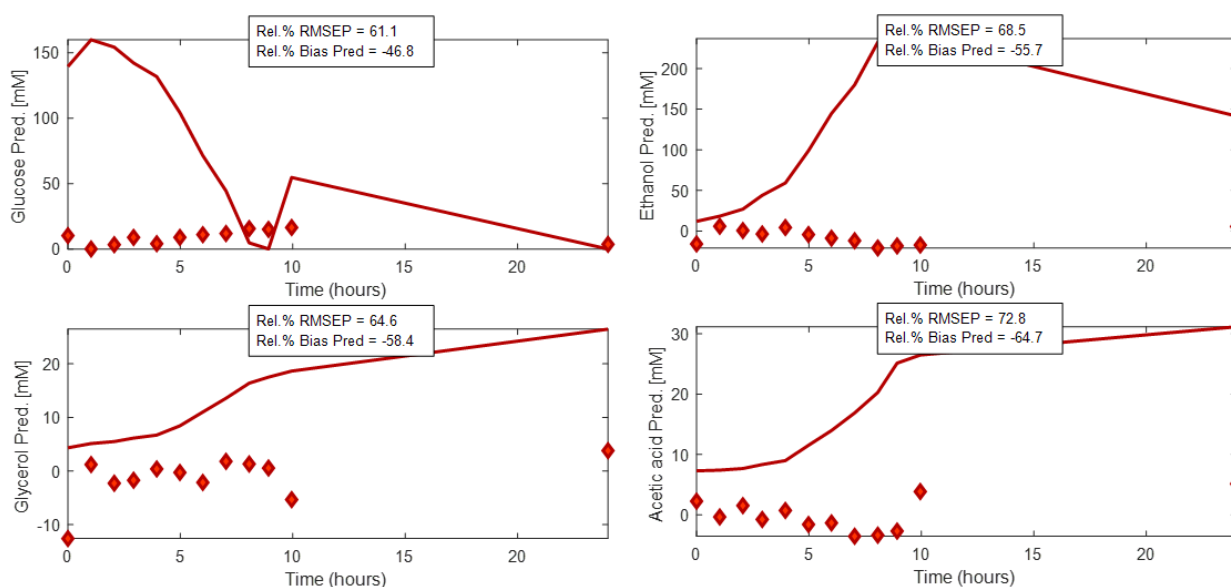


Figure S2.4: Model 3 (defined in Table 2.3 in the main text), built on offline ATR data applied to the flow-through ATR data, acquired on-line. HPLC data (solid line) and predicted data (filled diamonds) as a function of fermentation time. Note that, the HPLC data is presented as a continuous line to distinguish it from the predictions. However, it is based on the connection of the distinct data points and does not show the actual consumption / production profile between 10 and 24 h. Samples were collected hourly over the first 10 hours and a last sample after 24 h.

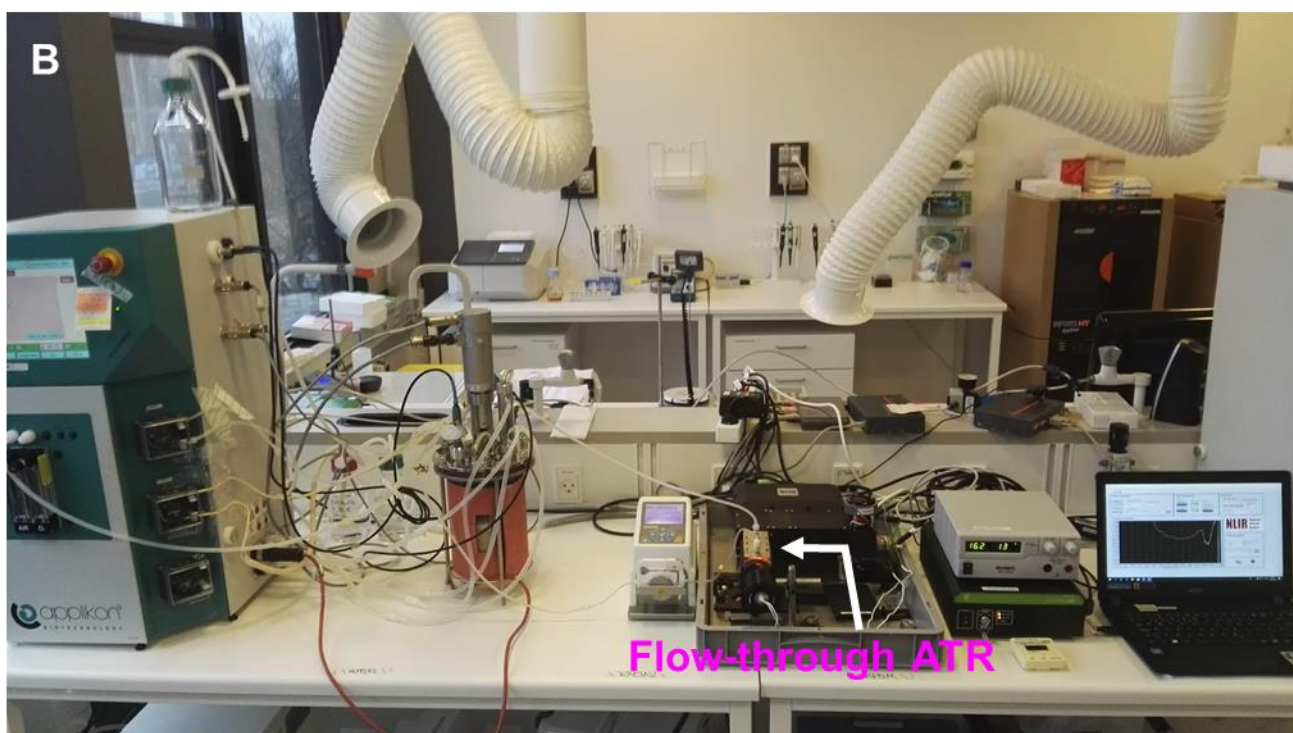
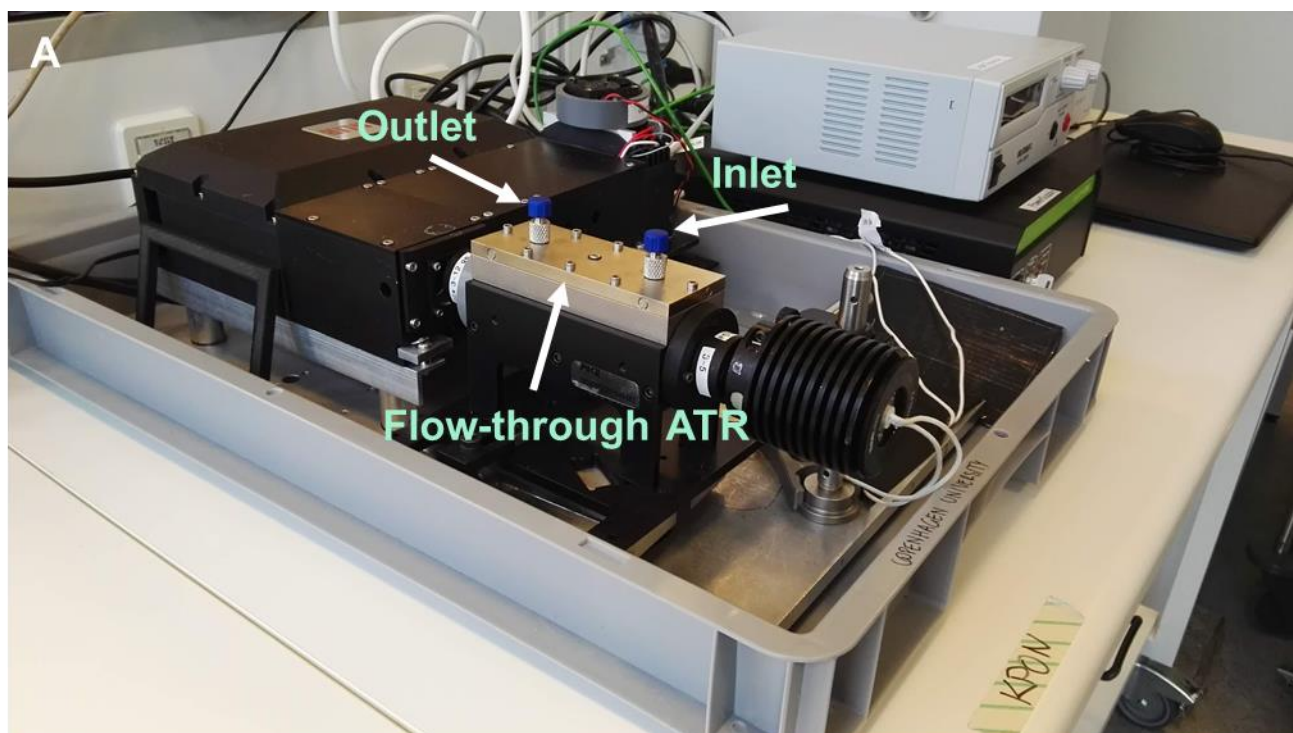


Figure S2.5: A) NLIR instrument equipped with the flow-through ATR element. B) Fermentation with continuous on-line collection of IR-spectra by means of the NLIR instrument equipped with the flow-through ATR-element.

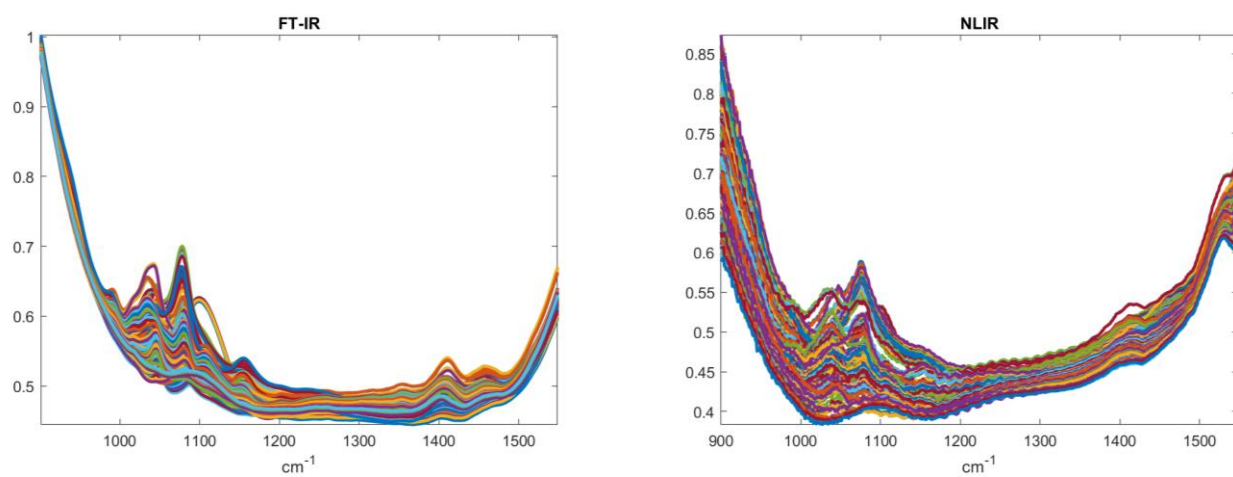


Figure S2.6: Comparison of all fermentation spectra collected with the FT-IR instrument (left) and the NLIR instrument (right)

2.6 References

- [1] Peter Larkin, "Infrared and Raman Spectroscopy: Principles and Spectral Interpretation," *Infrared and Raman Spectroscopy: Principles and Spectral Interpretation*. pp. 73–115, 2011.
- [2] D. W. Sun, "Infrared Spectroscopy for Food Quality Analysis and Control," *Infrared Spectroscopy for Food Quality Analysis and Control*. 2009.
- [3] D. L. Pavia, G. M. Lampman, and G. S. Kriz, *Introduction to spectroscopy: a guide for students of organic chemistry*. Harcourt College Publishers, 2001.
- [4] B. K. Lavine, "Chemometrics," *Anal. Chem.*, vol. 72, no. 12, pp. 91–98, Jun. 2000.
- [5] K. R. Beebe and B. R. Kowalski, "An Introduction to Multivariate Calibration and Analysis," *Anal. Chem.*, vol. 59, no. 17, 1987.
- [6] D. Steele, "Multivariate calibration," *Spectrochim. Acta Part A Mol. Spectrosc.*, vol. 46, no. 10, p. 1541, Jan. 1990.
- [7] M. Sandor, F. Rüdinger, R. Bienert, C. Grimm, D. Solle, and T. Scheper, "Comparative study of non-invasive monitoring via infrared spectroscopy for mammalian cell cultivations," *J. Biotechnol.*, vol. 168, no. 4, pp. 636–645, 2013.
- [8] A. Subramanian and L. Rodriguez-Saona, "Fourier Transform Infrared (FTIR) Spectroscopy," in *Infrared Spectroscopy for Food Quality Analysis and Control*, Elsevier, 2009, pp. 145–178.
- [9] U. Guide, "Installation and User Guide HATR Flow Cell , Jacketed Flow Cell , Heated Flow Cell for Horizontal ATR Accessories."
- [10] A. Bogomolov *et al.*, "Development and testing of mid-infrared sensors for in-line process monitoring in biotechnology," *Sensors Actuators, B Chem.*, vol. 221, pp. 1601–1610, 2015.
- [11] P. Fayolle, D. Picque, and G. Corrieu, "On-line monitoring of fermentation processes by a new remote dispersive middle-infrared spectrometer," *Food Control*, vol. 11, no. 4, pp. 291–296, 2000.
- [12] D. Landgrebe *et al.*, "On-line infrared spectroscopy for bioprocess monitoring," *Appl. Microbiol. Biotechnol.*, vol. 88, no. 1, pp. 11–22, 2010.
- [13] J. S. Dam, L. Høgstædt, C. Pedersen, P. Tidemand-Lichtenberg, and H. V. Andersen, "Mid-infrared upconversion spectroscopy," *J. Opt. Soc. Am. B*, vol. 33, no. 11, p. D28, 2016.
- [14] A. Barh, C. Pedersen, and P. Tidemand-Lichtenberg, "Ultra-broadband mid-wave-IR upconversion detection," *Opt. Lett.*, vol. 42, no. 8, p. 1504, 2017.
- [15] H. Kornmann, S. Valentinotti, P. Duboc, I. Marison, and U. von Stockar, "Monitoring and control of *Gluconacetobacter xylinus* fed-batch cultures using in situ mid-IR spectroscopy," *J. Biotechnol.*, vol. 113, no. 1–3, pp. 231–245, Sep. 2004.
- [16] C. Koch, A. E. Posch, H. C. Goicoechea, C. Herwig, and B. Lendl, "Multi-analyte quantification in bioprocesses by Fourier-transform-infrared spectroscopy by partial least squares regression and multivariate curve resolution," *Anal. Chim. Acta*, vol. 807, pp. 103–110, Jan. 2014.

- [17] J. Dahlbacka, J. Weegar, N. von Weymar, and K. Fagervik, "On-line measurement of the substrate concentrations in *Pichia pastoris* fermentations using FT-IR/ATR," *Biotechnol. Lett.*, vol. 34, no. 6, pp. 1009–1017, Jun. 2012.
- [18] N. D. Lourenço, J. A. Lopes, C. F. Almeida, M. C. Sarraguça, and H. M. Pinheiro, "Bioreactor monitoring with spectroscopy and chemometrics: a review," *Anal. Bioanal. Chem.*, vol. 404, no. 4, pp. 1211–1237, Sep. 2012.
- [19] P. D. *et al.*, "Real-time analyte monitoring of a fungal fermentation, at pilot scale, using in situ mid-infrared spectroscopy," *Bioprocess Biosyst. Eng.*, vol. 24, no. 1, pp. 13–24, Aug. 2001.
- [20] T. Genkawa, M. Watari, T. Nishii, and Y. Ozaki, "Development of a Near-Infrared/Mid-Infrared Dual-Region Spectrometer for Online Process Analysis," *Appl. Spectrosc.*, vol. 66, no. 7, pp. 773–781, Jul. 2012.
- [21] S. Sakhamuri, J. Bober, J. Irudayaraj, and A. Demirci, "Simultaneous determination of multiple components in lactic acid fermentation using FT-MIR, NIR and FT-Raman spectroscopic techniques," *Process Biochem.*, vol. 37, pp. 371–378, 2001.
- [22] N. Petersen, P. Ödman, A. E. Cervera Padrell, S. Stocks, A. E. Lantz, and K. V. Gernaey, "In situ near infrared spectroscopy for analyte-specific monitoring of glucose and ammonium in *Streptomyces coelicolor* fermentations," *Biotechnol. Prog.*, vol. 26, no. 1, pp. 263–271, 2010.
- [23] C. E. Eskildsen, M. A. Rasmussen, S. B. Engelsen, L. B. Larsen, N. A. Poulsen, and T. Skov, "Quantification of individual fatty acids in bovine milk by infrared spectroscopy and chemometrics: Understanding predictions of highly collinear reference variables," *J. Dairy Sci.*, vol. 97, no. 12, pp. 7940–7951, Dec. 2014.
- [24] D. T. Berhe, C. E. Eskildsen, R. Lametsch, M. S. Hviid, F. van den Berg, and S. B. Engelsen, "Prediction of total fatty acid parameters and individual fatty acids in pork backfat using Raman spectroscopy and chemometrics: Understanding the cage of covariance between highly correlated fat parameters," *Meat Sci.*, vol. 111, pp. 18–26, 2016.
- [25] Z. Liu, K. E. J. Tyo, J. L. Martínez, D. Petranovic, and J. Nielsen, "Different expression systems for production of recombinant proteins in *Saccharomyces cerevisiae*," *Biotechnol. Bioeng.*, 2012.
- [26] B. Sonnleitner and O. Käppeli, "Growth of *Saccharomyces cerevisiae* is controlled by its limited respiratory capacity: Formulation and verification of a hypothesis," *Biotechnol. Bioeng.*, vol. 28, no. 6, pp. 927–937, Jun. 1986.
- [27] R. von Schalien, K. Fagervik, B. Saxén, K. Ringbom, and M. Rydström, "Adaptive on-line model for aerobic *Saccharomyces cerevisiae* fermentation," *Biotechnol. Bioeng.*, vol. 48, no. 6, pp. 631–638, Dec. 1995.
- [28] K. Buse, F. Kühnemann, S. Wolf, J. Kiessling, M. Kunz, and G. Popko, "Upconversion-enabled array spectrometer for the mid-infrared, featuring kilohertz spectra acquisition rates," *Opt. Express*, vol. 25, no. 13, p. 14504, 2017.
- [29] M. Sjöström, S. Wold, W. Lindberg, J.-Å. Persson, and H. Martens, "A multivariate calibration problem in analytical chemistry solved by partial least-squares models in latent variables," *Anal. Chim. Acta*, vol. 150, no. 150, pp. 61–70, 1983.

- [30] P. Geladi and B. R. Kowalski, "Partial least-squares regression: a tutorial," *Anal. Chim. Acta*, vol. 185, no. 185, pp. 1–17, 1986.
- [31] A. Savitzky and M. J. E. Golay, "Smoothing and Differentiation of Data by Simplified Least Squares Procedures," *Anal. Chem.*, vol. 36, no. 8, pp. 1627–1639, Jul. 1964.
- [32] R. J. Barnes, M. S. Dhanoa, and S. J. Lister, "Standard Normal Variate Transformation and De-Trending of Near-Infrared Diffuse Reflectance Spectra," *Appl. Spectrosc.*, vol. 43, no. 5, pp. 772–777, Jul. 1989.
- [33] Åsmund Rinnan, F. van den Berg, and S. B. Engelsen, "Review of the most common pre-processing techniques for near-infrared spectra," *TrAC Trends Anal. Chem.*, vol. 28, no. 10, pp. 1201–1222, Nov. 2009.
- [34] S. Wold, K. Esbensen, and P. Geladi, "Principal component analysis," *Chemom. Intell. Lab. Syst.*, vol. 2, no. 1–3, pp. 37–52, Aug. 1987.
- [35] K. Pearson, "LIII. On lines and planes of closest fit to systems of points in space," *London, Edinburgh, Dublin Philos. Mag. J. Sci.*, vol. 2, no. 11, pp. 559–572, Nov. 1901.
- [36] H. Kornmann, M. Rhiel, C. Cannizzaro, I. Marison, and U. Von Stockar, "Methodology for real-time, multianalyte monitoring of fermentations using an in-situ mid-infrared sensor," *Biotechnol. Bioeng.*, vol. 82, no. 6, pp. 702–709, 2003.
- [37] H. Wu *et al.*, "Real time monitoring of bioreactor mAb IgG3 cell culture process dynamics via Fourier transform infrared spectroscopy: Implications for enabling cell culture process analytical technology," *Front. Chem. Sci. Eng.*, vol. 9, no. 3, pp. 386–406, 2015.
- [38] J. Schenk, I. W. Marison, and U. von Stockar, "Simplified Fourier-transform mid-infrared spectroscopy calibration based on a spectra library for the on-line monitoring of bioprocesses," *Anal. Chim. Acta*, vol. 591, no. 1 SPEC. ISS., pp. 132–140, 2007.
- [39] P. Roychoudhury, L. M. Harvey, and B. McNeil, "At-line monitoring of ammonium, glucose, methyl oleate and biomass in a complex antibiotic fermentation process using attenuated total reflectance-mid-infrared (ATR-MIR) spectroscopy," *Anal. Chim. Acta*, vol. 561, no. 1–2, pp. 218–224, 2006.
- [40] J.-J. Max and C. Chapados, "Aqueous ammonia and ammonium chloride hydrates: Principal infrared spectra," *J. Mol. Struct.*, vol. 1046, pp. 124–135, Aug. 2013.
- [41] E. L. Wagner and D. F. Hornig, "The Vibrational Spectra of Molecules and Complex Ions in Crystals IV. Ammonium Bromide and Deutero-Ammonium Bromide," *J. Chem. Phys.*, vol. 18, no. 3, pp. 305–312, Mar. 1950.
- [42] M. Klähn *et al.*, "IR Spectra of Phosphate Ions in Aqueous Solution: Predictions of a DFT/MM Approach Compared with Observations," *J. Phys. Chem. A*, vol. 108, no. 29, pp. 6186–6194, Jul. 2004.
- [43] W. W. Rudolph, "Raman- and infrared-spectroscopic investigations of dilute aqueous phosphoric acid solutions," *Dalt. Trans.*, vol. 39, no. 40, p. 9642, 2010.
- [44] B. R. Singh, "Basic Aspects of the Technique and Applications of Infrared Spectroscopy of Peptides and Proteins," pp. 2–37, 2009.

- [45] R. G. Brereton, *2 Experimental Design*, vol. 8. 2003.
- [46] S. Sivakesava, J. Irudayaraj, and A. Demirci, "Monitoring a bioprocess for ethanol production using FT-MIR and FT-Raman spectroscopy," *J. Ind. Microbiol. Biotechnol.*, vol. 26, no. 4, pp. 185–190, 2001.
- [47] A. Bogomolov, "Diagonal designs for a multi-component calibration experiment," *Anal. Chim. Acta*, vol. 951, pp. 46–57, 2017.
- [48] O. M. Kvalheim, "Interpretation of partial least squares regression models by means of target projection and selectivity ratio plots," *J. Chemom.*, vol. 24, no. 7–8, pp. 496–504, Jul. 2010.

Chapter 3

Monitoring of microbial growth and morphology dynamics by imaging and image analysis

Preface

The following chapter introduces the exciting possibility of monitoring a yeast fermentation process by means of recent advances in automatic imaging and image analysis. It is exemplarily employed on the oCelloScope technology. The instrument and the complementary UniExplorer and ParticleTech software were used in order to study the potential of available automatic imaging and image analysis algorithms with respect to detection of yeast growth and morphology dynamics in the course of a batch fermentation with yeast. The oCelloScope technology has successfully been applied to crystallization processes monitoring as well as to microbial and cancer research, conducted on microtiter plates. However, it has not been used to literally look inside a lab-scale yeast fermentation via an on-line recirculation loop as it was done in this work. In order to enable the investigation of yeast morphology dynamics with respect to single cells, budding cells and cell-clusters over the fermentation time, the so-called bright spot feature was developed. The algorithm enables the automatic differentiation between the cell-objects of interest (single cells, budding cells and cell clusters) and hence allowed to yield first trend with respect to the morphology dynamics under study. The first morphology dynamics obtained can be considered as an effective demonstration of the potential of this approach but cannot be taken into account for a final conclusion regarding this research. Note that, the oCelloScope technology progressed in parallel with this work. For the very first time, it could be connected continuously to a lab-scale fermentation process by means of a prototype flow-through-cell. The initial idea of using this instrument as a continuous monitoring device could finally be transferred into practice by support of ParticleTech Aps. I would like to acknowledge Eric Spillum (BioSense Solutions Aps) for developing the bright spot feature as well as Trine Aabo Anderson and Louise la Cour Freiesleben (ParticleTech Aps) for enabling the first monitoring experiment conducted with the oCelloScope instrument on a yeast lab-scale batch fermentation.

3.1 Introduction

In any fermentation process, the quantification of microbial growth or biomass, respectively, is the most evident parameter to evaluate the progress of the fermentation. The microbial biomass is the producing core of any biotechnological product and frequently, the biomass is the product itself. Next to substrate and product concentration levels, the monitoring of the biomass concentration gives crucial information about the process performance. However, the quantification of this essential

parameter is mostly limited to off-line analysis and common detection systems are challenged especially by the very early or late stages of the fermentation process, where minor or, respectively, very high biomass concentrations are present.

The most established quantification method used in biotechnology to detect microbial growth in liquid samples is the measurement of optical density (OD). In OD measurements, light at a wavelength of (usually) 600 nm is transmitted through the sample and scattered by the cells in suspension. Thus, the more cells are present, the more light is scattered and the less light reaches the detector. The higher the biomass concentration, the lower the intensity of the detected light and the higher the resulting OD value. The conventionally used wavelength of 600 nm presents a good compromise between availability of filters on the market and both, sensitivity and selectivity of the measurement. Higher wavelengths generate higher scattering than lower wavelengths and absorption due to biomolecules present in the sample (*e.g.* due to proteins or nucleic acids at 230 nm and 260 – 280 nm, respectively) is avoided. However, a reasonable number of cells must be present to cause first changes and the very early stages of growth cannot be detected by OD-measurements. The same holds for high cell concentrations reached at the end of a fermentation process. Common spectrophotometers are usually limited to detect OD values between 0.1 and 1 accounting for approximately 0.1 and 1 g/l cell dry mass respectively, while in most fermentation processes a cell dry weight concentration of more than 30 g/l can be expected, with cell concentration levels going up to 100 g/l in large scale high cell density processes. Thus, in order to rely on OD measurements at high biomass concentrations, samples must be properly diluted making this technique not applicable on-line, when considering that no advanced automated sampling unit is available. Another draw-back of this method is that OD measurements are highly dependent on the optical system and its geometry in use (*e.g.* area and sensitivity of the detector, distance between sample and detector, etc.). Accordingly, spectrophotometers with different optical configurations result in different OD values for the same sample. Moreover, cells in different physiological states (*e.g.* dormant, vital, dead) cannot be distinguished and basically any microbial cell independent of its biological state and nature, in the worst case a contamination, contributes to the OD value. Regardless these limitations, most biotechnological applications still predominantly rely on OD measurements to quantify microbial growth. The measurement is easy and fast to perform and must be considered as a standard method across biotechnology.

The constant need for accurate, sensitive and specific detection technologies for biological samples has also driven the area of biomass detection. Recent advances in microscopy, imaging and image analysis have opened a visual window to literally observe and evaluate growth and cellular events based on parameters associated with the visual observation of the cells. The imaging of living cells has yielded tremendous insights into cellular growth, functions and responses to environmental changes. The degree of automation in (light) microscopy has achieved great advances and the technique has grown into an efficient tool integrating imaging and image analysis. Bright-field [1] and fluorescence microscopy [2] are often used to access information about the cell size, shape,

position and motility of living cells [3], [4], [5]. Challenges arising from those techniques, such as a small field of view introducing poor counting statistics and the visualization of so called 'phase objects' - objects appearing transparent on the image - are solved by phase contrast [6] and confocal [7] microscopy. However, the latter techniques are expensive, prone to a dedicated environment, require a long acquisition time and have the disadvantage that improper settings can cause significant artefacts in the images [8].

In general, bright field microscopy is the simplest available microscopy technique and the latest application solutions offer the possibility to literally gain new insight into the bioreactor. In bright field microscopy, white light is transmitted through the sample. Cells in the sample absorb part of the light and thus, they can be distinguished from the brighter background on the images. The oCelloScope instrument (BioSense Solutions ApS) is a new, compact and portable solution based on bright field digital [9], time-lapse [10] microscopy. The imaging system consists of a digital camera, an illumination unit and a system of lenses integrated in a way that the optical axis is tilted 6.25° relative to the horizontal plane of the stage. Due to this tilt, scanning of volumes and extraction of phase information are possible defining in particular the scope of this technology. The system consists of the oCelloScope instrument and the UniExplorer software for instrument control and image analysis. Acquisition of images is simplified by means of integrated algorithms enabling automatic adjustment of focus and illumination level. On- and off-line image processing is facilitated by additional advanced algorithms accessing microbial growth and morphological features associated with the objects identified simultaneously. Moreover, the system is based on an open technology platform that can be scaled to a wide array of applications. The open technology enabled the implementation of the so-called bright spot feature into the software, developed within this work. Hence, the software can be adjusted and tailored towards the dedicated needs of a particular application. The analysis tools provided by the UniExplorer software have been shown to provide considerable advantages in several research fields including microbiology [11], medicinal chemistry [12], pharmaceutical biotechnology [13] and basic cancer research [14]. Moreover, the technology has found a solid application area in monitoring of bacterial growth, or respectively, growth inhibition [15] and changes in microbial morphology [16]. Thus far it has not been applied as a monitoring device for fermentation processes, as investigated in this chapter.

However, the implemented tools for automated imaging and image analysis suggest this system as a novel promising monitoring technique within the scope of fermentation monitoring. Acquisition and analysis of an image is performed within a minute. Moreover, the simultaneous detection of microbial growth and morphological features make this system in particular attractive. In yeast cultivations, information about the cell size and cell size distribution, respectively, has been shown to be correlated with the cell viability (dead/alive, osmotically stressed) [17], [18] and the growth rate of the culture [19]. Furthermore, the cell size was recently correlated to the accumulation of an internal product (fatty acids) in microalgae [20]. Consequently, image analysis seems to be a promising tool for obtaining a snapshot of the physiological state of a yeast culture by assessing its morphological parameters. Ultimately, a snapshot of the culture can be used to assess the most

crucial parameters as microbial growth, substrate and product levels at a specific time point, in the end enabling novel image based control strategies of the reactor. The possibility to have a quick look into the fermentation process provides a novel and interesting aspect. Some sources of failure as *e.g.* contamination can literally be seen from the beginning. Hence, in case of a contamination observed in the very early stages of the process, the process can be stopped without wasting more time and money on it.

Within this chapter, a first trial of image based monitoring of a lab-scale yeast fermentation process is demonstrated. So far, the oCelloScope has supported several types of sample containers including microscope slides and microtiter plates up to 96 wells. For the very first time in the current study, the oCelloScope instrument was continuously connected to the fermenter via a prototype flow-through-cell. The continuous on-line set-up enabled frequent, automated imaging over a 10 h batch yeast fermentation. For this application aiming at growth detection complemented with the evaluation of morphological trends, the so called bright-spot feature was developed. The algorithm facilitates the automatic differentiation between single cells, budding cells and cell-clusters. By analyzing images over the cultivation time, the distribution dynamics between single cells, budding cells and cell-clusters were investigated, aiming at establishing correlations between morphological features and process performance. First time trends of the yeast morphology are discussed in detail.

The following chapter examines and discusses the potential of recent advances in automatic imaging and image analysis exemplarily demonstrated with the oCelloscope instrument. In the future, the combined, automated method of imaging and tailored image analysis may become a totally novel control strategy of fermentation processes.

3.2 Materials and Methods

3.2.1 The oCelloScope detection principle

The basic principle behind the oCelloScope detection system (BioSense Solutions ApS, <https://biosensesolutions.dk/>) is bright-field, digital time-lapse microscopy. The imaging unit is demonstrated in Figure 3.1.

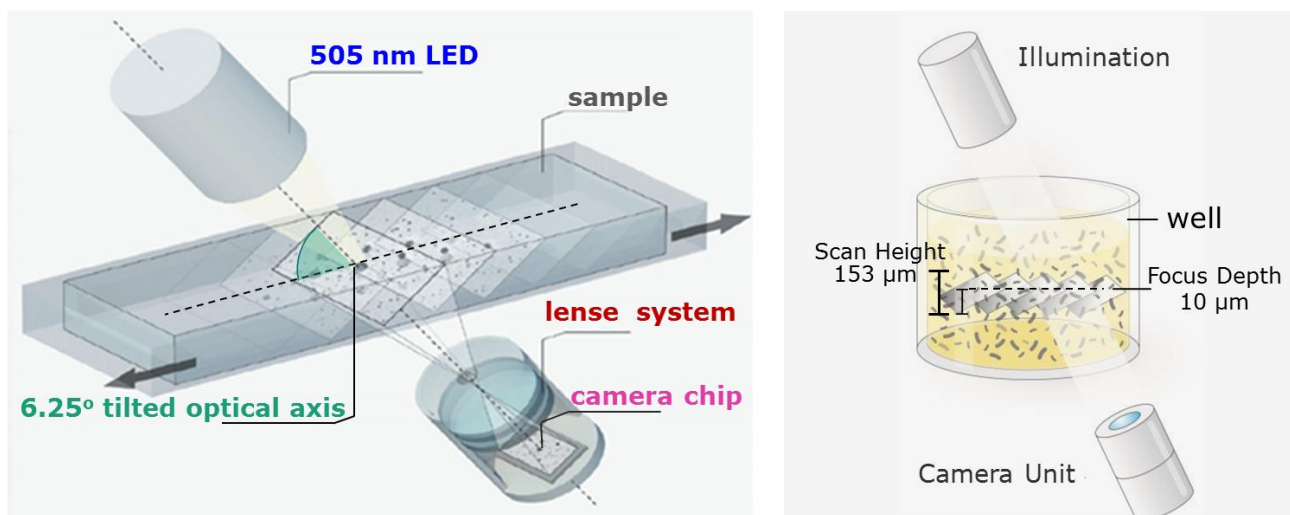


Figure 3.1: The oCelloScope detection principle (left) and a schematic presentation of a well containing a cell suspension demonstrating the scan height and focus depth of the imaging system (right). The figure is borrowed and modified with permission from BioSense Solutions ApS.

It consists of a 505 nm LED illumination unit, a proprietary lenses system and a 5-megapixel complementary metal oxide semiconductor (CMOS) camera chip (with length-by-height dimensions of 5.6 mm by 4 mm). The lenses and digital camera unit are designed to have a scan height of approximately 153 μm, a focus depth of approximately 10 μm, an absolute magnification factor of 4 and an optical resolution of 1.3 μm. With that, the system is comparable with a magnification factor of 200 in a standard light microscope facilitating the detection of objects with a size between 0.5 μm and 1 mm. Accordingly, it is suitable for the investigation of mammalian cells, yeast / fungi, bacteria and crystals in (semi) transparent substances. The optical axis of the imaging system is tilted 6.25° relative to the horizontal plane of the stage enabling scanning of volumes and extraction of phase information. When scanning through a liquid sample, a series of images (the number is defined in the software) is generated from which the final best focus image, as well as the out-of-focus images are generated, as presented in [Figure 3.2](#).

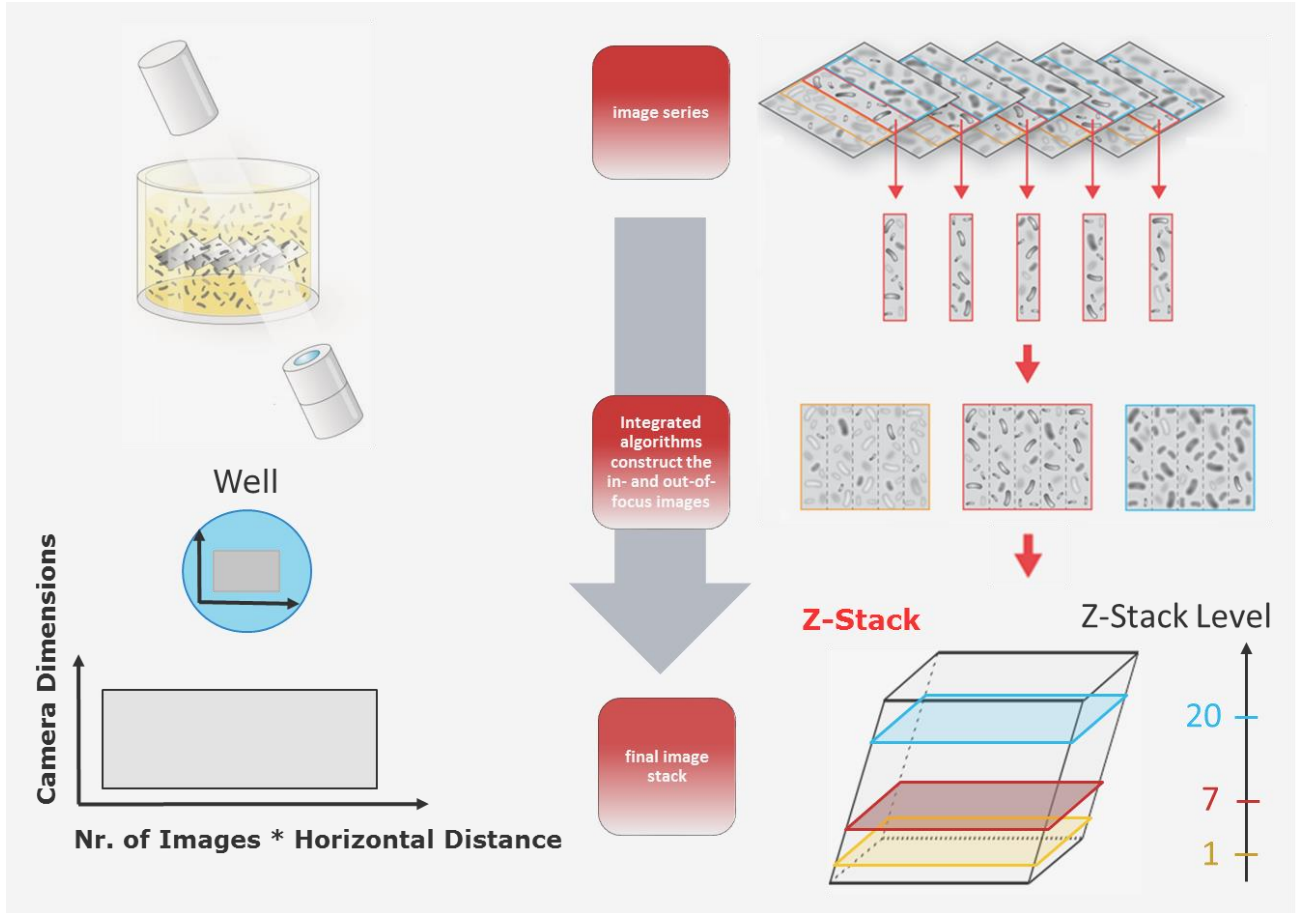


Figure 3.2: Schematic representation of the image dimension (left) and the creation of Z-Stack images based on the image series recorded. The Z-stack images represent a scan through the detection volume resulting in in-focus and out-of-focus images. The figure is borrowed and modified with permission from BioSense Solutions ApS

Due to the tilted image plane, the recorded images form an image stack ('Z-stack') representing the scan height of $153\ \mu\text{m}$ by 20 layers (in case of an image distance of $4.9\ \mu\text{m}$ as used for off-line analysis, (see 3.2.7 Off-line image acquisition), containing the images in- focus and out-of-focus across the detection volume. The detection volume of a sample has the shape of a parallelepiped and can be calculated according to equation (3.1)

$$V_{\text{sample}} = L_{\text{camera}} * H_{\text{camera}} * W_{\text{sample}} * \sin(6.25) * n_{\text{medium}} \quad \text{Equation (3.1)}$$

L_{camera} and H_{camera} are the camera chip dimensions downscaled with the magnification factor ($5.6\ \text{mm}/4 = 1.4\ \text{mm}$ and $4\ \text{mm}/4 = 1\ \text{mm}$). W_{sample} equals the scan length defined by the number of images acquired (set in the software) and the horizontal distance between them (minimal incremental step length). The latter is $7.5\ \mu\text{m}$ for the instrument in use. As the refractive index of the sample (aqueous medium) is higher than the refractive index of air, the object plane is displaced deeper into the medium and therefore, the final, truly observed volume of the sample must be multiplied by the refractive index of the medium, which can be approximated by the refractive index of water ($n_{\text{water}} = 1.34$, [15]).

Accordingly, the image area of each Z-stack layer is calculated by Equation (3.2).

$$A_{\text{image}} = H_{\text{camera}} * W_{\text{sample}}$$

Equation (3.2)

Another way to calculate the image area is by multiplying the image length and width in pixels, which can inherently be accessed by the image properties, by the size of a pixel. For the oCelloScope technology, the pixel size is approximately 0.325 μm (length and width of a pixel). It is obtained by the optical resolution (1.3 μm) of the oCelloScope downsampled with the magnification factor of 4. Note that, the resolution is fixed while the magnification factor varies slightly from oCelloScope to oCelloScope.

Image analysis was automatically performed on the best focus image, but all images might be considered for growth analysis and segmentation. The term segmentation describes the event of identifying the objects present on an image in their true shape and distinguishing them from the background. Within this study, a (cell-) object can be a single cell, a budding cell or a cell cluster of several sizes.

3.2.2 Detection of microbial growth

The oCelloScope provides several algorithms to constitute a growth curve based on the constructed best focus image of each time point. The following algorithms were applied in order to detect yeast growth based on image analysis:

- Background Corrected Absorption (BCA)

$$\text{BCA value} = \log_{10}(\sum \text{object pixel intensity})$$

The BCA algorithm corrects background intensities with respect to the first image acquired. This allows to obtain images with an even light distribution, which are used for calculating an intensity threshold. The threshold divides pixels into 'background' and 'objects'. This enables the BCA algorithm to detect small changes when new cells appear in the image. Growth curves are generated based on changes in 'objects' so that the effect of background intensities are significantly reduced.

- Total Absorption (TA)

TA value = $\log_{10}(\sum \text{pixel intensity})$

During microbial growth, the increasing number of objects will reduce the light transmission through the sample and the image will get progressively darker. A darker image is equivalent to a higher TA value. Sensitivity is limited compared to the BCA algorithm as growth and cell concentration need to be quite considerable before affecting the light transmitted through the sample.

The BCA and TA algorithm are both based on the same principle as an OD measurement. The pixel intensity can be regarded as being equal to the absorbance. Hence, both algorithms are based on absorbance, with the TA algorithm being equal to an OD measurement, while the BCA algorithm can be regarded as an OD measurement with higher sensitivity.

- Segmentation and Extraction of Surface Area (SESA)

SESA value = $\log_{10}(\sum \text{object covered area})$

The SESA algorithm identifies all the objects segmented in the best focus image based on their contrast against the background and calculates the total surface area covered by these objects. When more than 20% of the total image area is covered by objects, the accuracy of the SESA algorithm starts to decline.

The growth data resulting from each algorithm is reported as arbitrary units and must be correlated with an external measurement in order to set the value in a relevant dimension. The normalized version of each algorithm is also available subtracting the value obtained from the first image from all the following images. The normalized growth values were considered within this study.

3.2.3 Morphological descriptors

Several segmentation parameters describing the identified objects quantitatively are available in the software. Three of the implemented software descriptors, together with the bright spot feature developed in relation to this work, were chosen as morphological descriptors.

- Area: The area parameter measures the total number of pixels covered by an object. The area can *e.g.* be used to monitor changes in the object size over time or to discriminate the segmentation of objects based on size exclusion. It is not affected by the object shape, in other words, objects with an identical area value may have different shapes.
- Perimeter: The perimeter value is computed based on the length of the object border in pixels. Accordingly, large objects result in a higher perimeter value than small objects. However, the shape of an object affects the perimeter value. Based on the object shape, two objects may have the same area but result in different perimeter values. A circular object will

possess a lower perimeter value than any other shape. Thus, the perimeter value may be used to distinguish between objects with equivalent areas but different shapes.

- **Circularity:** The circularity descriptor measures how similar the object shape is to a circle independently of the object size. The circularity value is calculated as the ratio between the perimeter of a circle with the same area as the object and the perimeter of the object. The value varies between 0 and 1. The closer to 1, the better the object shape represents a circle. Due to technical irregularities in the software, the circularity value of a circular object can be marginally larger than 1.
- **Bright Spots:** The bright spot descriptor was developed with respect to the application on a yeast fermentation, as yeast cells appear as bright spots on the image. The bright spot feature counts the number of bright spots associated to an object. With respect to this study, an object can be a single cell (one bright spot per object), a budding cell (two bright spots per object) and a cell-cluster containing several cells (three and more than three bright spots per object). This is demonstrated in [Figure 3.3](#). As the algorithm used for segmentation is based on a gradient threshold, objects with zero bright spots may be identified due to an uneven light distribution in the background. Exemplarily, objects identified with zero bright spots can be found in the supplementary material, Figure S3.1. Moreover, [Figure 3.3](#) demonstrates the challenge to correctly segment larger cell clusters, as they may be the result of adjacent objects (with touching border pixels) However, the algorithm works especially well for single and budding cells and cell clusters of approximately up to 5 cells.

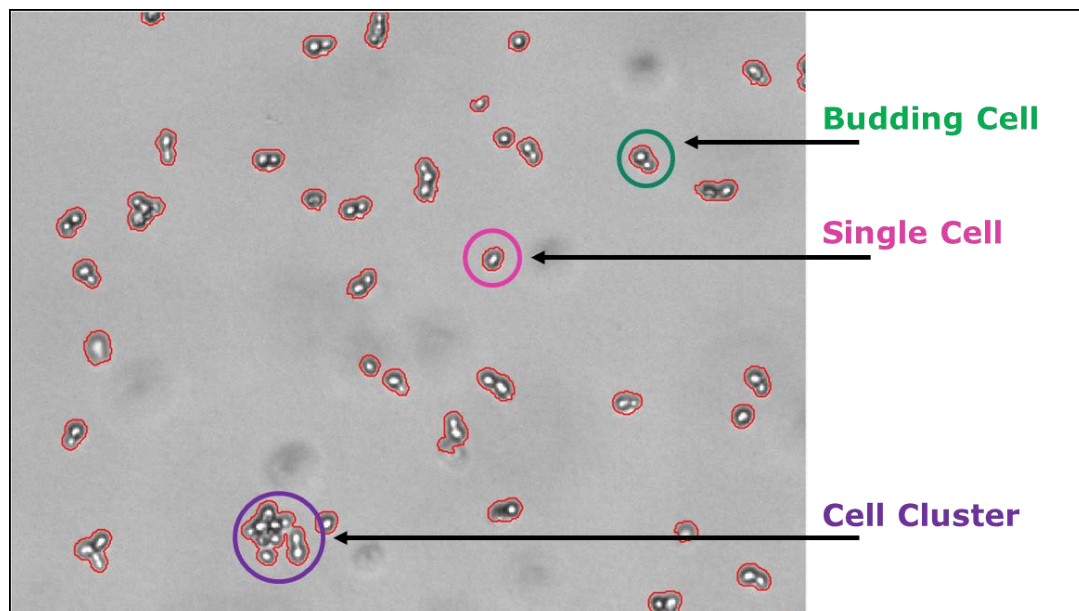


Figure 3.3: Image of a yeast cell culture possessing an OD value of approximately 0.1. Yeast cells appear as a bright spot surrounded by a darker border. The bright spot feature counts the number of bright spots per object. As indicated in the figure, one bright spot represents a single yeast cell, two bright spots represent a budding yeast cell and three and more than three bright spots represent a cluster of yeast cells.

The morphological parameters described were chosen in order to be able to get a first insight into the morphology dynamics over the fermentation course. In particular the bright spot feature was expected to describe the evolution of cell-objects over the fermentation time. The parameters area, perimeter and circularity will help to evaluate the consistency of the data. It is expected that the larger the area of an object, the larger the perimeter and the higher the number of bright spots. The circularity is expected to be close to 1 for a single cell and lower for any other cell object present. Note that, two different oCelloScope instruments and software versions were used due to its availability. The essential algorithms are the same in both software (UniExplorer and ParticleTech software). However, some small differences regarding the application of the segmentation algorithms were practically present, as mentioned were relevant in the coming sections.

3.2.4 Yeast fed-batch fermentation process

A detailed description of the chemicals and equipment used, medium composition as well as performed HPLC analysis can be found in 2.2.1.1 Yeast fermentation processes and 2.2.3.1 Glucose and metabolite quantification. A short summary of the fermentation process with its most crucial settings is described in the following.

The laboratory yeast strain CENPK 113-7D was cultivated in YPD medium (yeast extract (Y), peptone (P), dextrose (glucose, (D) dissolved in water)) for 10 h. The fermentation was performed in a volume of 2 L, controlled at a pH of 6 and a temperature of 30 °C. The aeration rate was set to 1 vvm and the agitation rate was controlled at a stirrer speed of 800 rpm. The dissolved oxygen tension (DOT) stayed above 30 % of saturation all through the process indicating that no oxygen limitation occurred during the cultivations. The fermentation vessel was equipped with two standard sampling ports. The first sampling port was used for continuous recirculation of the fermentation broth via the oCelloScope flow-through-cell. This sampling port was equipped with a 20 µm stainless steel filter cap(Filter screen, Applikon, The Netherlands) in order to prevent the sampling of air. The second sampling port was used for manual sampling via a sterile syringe in order to yield samples for reference analysis. Samples taken manually were analyzed by HPLC for glucose and metabolite concentration levels as well as by OD₆₀₀ measurements following microbial growth. Additionally, an advanced prototype of the NxPAS gas analyzer system was connected to obtain on-line data of the off-gas (CO₂). The NxPAS gas analyzer (<http://www.nxpas.com>) is based on a novel photoacoustic detection principle enabling the detection of CO₂, O₂ and several other molecules through IR spectra analysis. No more details of the system can be given within this work. Based on the online CO₂ data, the best suited time point for feeding glucose could be decided. 100 ml of glucose solution (400 g/l) was added by means of a pump (flow 10 ml/min, Ismatec Reglo ICC, Ismatec, Germany) after ca. 6 h (380 min) in order to prolong the exponential growth phase.

3.2.5 Investigating image acquisition and analysis – preliminary experiments

Growth experiments conducted on a 96 well-plate as well as in a lab-scale fermentation with equal conditions to the fermentation described above were used to study this technique with respect to growth analysis and morphological description of the cells. A growth experiment, performed directly

inside the oCelloScope on a 96 well-plate was used to get a first idea of the performance of the designed bright-spot feature and growth analysis in general. In case of the lab-scale fermentation, images of fermentation samples were recorded off-line on a 96 well-plate. With the help of these experiments, the practical settings for off-line imaging (illumination level and focus) via a 96 well-plate (Costar, flat-bottom, Germany) were defined. Moreover, it was found that a sample volume of 100 μ l was appropriate for imaging and image analysis. It was noticed that an OD value of higher than 2 resulted in a poor image quality with respect to segmentation analysis as the recorded images were too crowded with cells. This limited the identification of individual objects as the borders of objects may not clearly be defined due to directly adjacent and overlapping cells. Hence, an OD value of approximately 0.1 was decided to produce a good image quality with cell-objects showing a reasonable distance from each other on the images, enabling a reliable segmentation with enough objects to generate statistically meaningful data. As the sample preparation (dilution and mixing) might affect the cell morphology with respect to cluster size, the influence of mixing power and dilution medium on the cell morphology was studied as well. It was found that both, the dilution medium and the mixing power affect the cell morphology dynamics slightly. However, the destruction of cell clusters due to an improper dilution medium or exaggerated mixing was found to be minor. Saline and the fermentation medium itself (without glucose, referred to as YP medium) were considered as proper dilution media. Saline was chosen over YP medium due to practical reasons. YP medium is leading to foaming whereas saline is not. Hence, samples diluted in saline could be pipetted smoothly into the microliter plate without the need to exclude air bubbles manually after the transfer process.

3.2.6 On-line image acquisition

An oCelloScope operated by the ParticleTech Software was connected on-line in such a way that images could be acquired automatically every 10 minutes. Recirculation of the fermentation broth was facilitated by means of a peristaltic pump (Longer, USA) via a prototype of the oCelloScope flow-through-cell. The flow-cell was 3D printed in a transparent hard plast material as a demo version. The dimensions of the flow through chamber were 6 mm x 20 mm x 0.8 mm, containing a sample volume of 96 μ l. The pump controlled by the ParticleTech software recirculated the fermentation broth from a sampling port equipped with a 20 μ m stainless steel filter cap (sampling scree, Applikon, The Netherlands) inside the bioreactor through the flow-cell inside the oCelloScope and back to the fermenter (using a separated inlet). For image acquisition, the pump stopped automatically and an image was acquired after a defined settling time of 5 seconds. When the image acquisition finished, the pump restarted and was in operation until the next image was to be collected. The dead volume inside the sampling pipe and the tubing until the flow-through-cell was approximately 7.5 ml. The pump was operated at a flow rate of 1.3 ml/min and positioned after the flow-through-cell. In this way, cell (-cluster) damage due to the pump activity was avoided and an exchange of the volume inside the flow-cell with fresh fermentation broth was guaranteed approximately every 8th minute. Note that, a higher flow rate might be more suitable in order to avoid the sticking of cells to the inner surface of the flow-cell and would facilitate a higher exchange rate of volume. However, in this

experiment, the flow rate was limited since a microfluidic flow-through-biosensor (described in detail in chapter 1, section 1.2.1 Glucose determination using a biosensor flow-through cell) was connected to the same recirculation loop (located after the oCelloScope). The biosensor data is not part of this chapter. However, the inclusion of the biosensor into the recirculation loop limited the flow rate to a maximal value of 1.3 ml/min. In the on-line set-up facilitated by means of the prototype flow-through-cell, the focus and the illumination level was set automatically. The image size was defined by an image distance of 45 μm and a number of 100 images, collected in 2 rows, representing a scanned volume of 9.3 nl and a final image area of 2.1 mm^2 (see Equation 3.1 and 3.2).

3.2.7 Off-line image acquisition

Fermentation samples, manually withdrawn every 40 minutes, were analyzed off-line via an oCelloScope operated by the UniExplorer software. Images of the fermentation samples were acquired on both, original and diluted fermentation samples. After the preliminary experiments, a reasonable set-up for imaging and image analysis conducted off-line on a 96 well plate was defined as the following: The OD value of each fermentation sample was determined and samples were diluted in sterile saline (9 g/l NaCl solution in purified water) to an OD value of approximately 0.1. Every cell sample was pipetted in a volume of 100 μl into the microtiter plate, producing duplicates in two separated wells. The focus was set to be around 2985 μm and the illumination level was set to 175. The image size was defined by an image distance of 4.9 μm and a number of 50 images, representing a scanned sample volume and a final image area of 2.3 nl and 0.53 mm^2 , respectively (see Equation 3.1 and 3.2). In this way, images with good quality aiming at proper image analysis were acquired and analyzed with respect to growth (undiluted samples) and changes in morphology (diluted samples).

3.2.8 Experimental set-up

Two instances of the oCelloScope instrument were used to follow the lab-scale fermentation on-line and off-line. The on-line set-up was operated by the ParticleTech software while the off-line set-up was operated by the Uniexplorer software. Both software versions apply essentially the same algorithms for microbial growth and morphology description. The ParticleTech software was developed with respect to cristal imaging, the Uniexplorer software was developed within the scope of cell investigations, hence the different versions.

A picture of the whole set-up including a representation of the different sampling devices for the oCelloScope instrument (96 well-plate vs. flow-through-cell) is shown in [Figure 3.4](#).

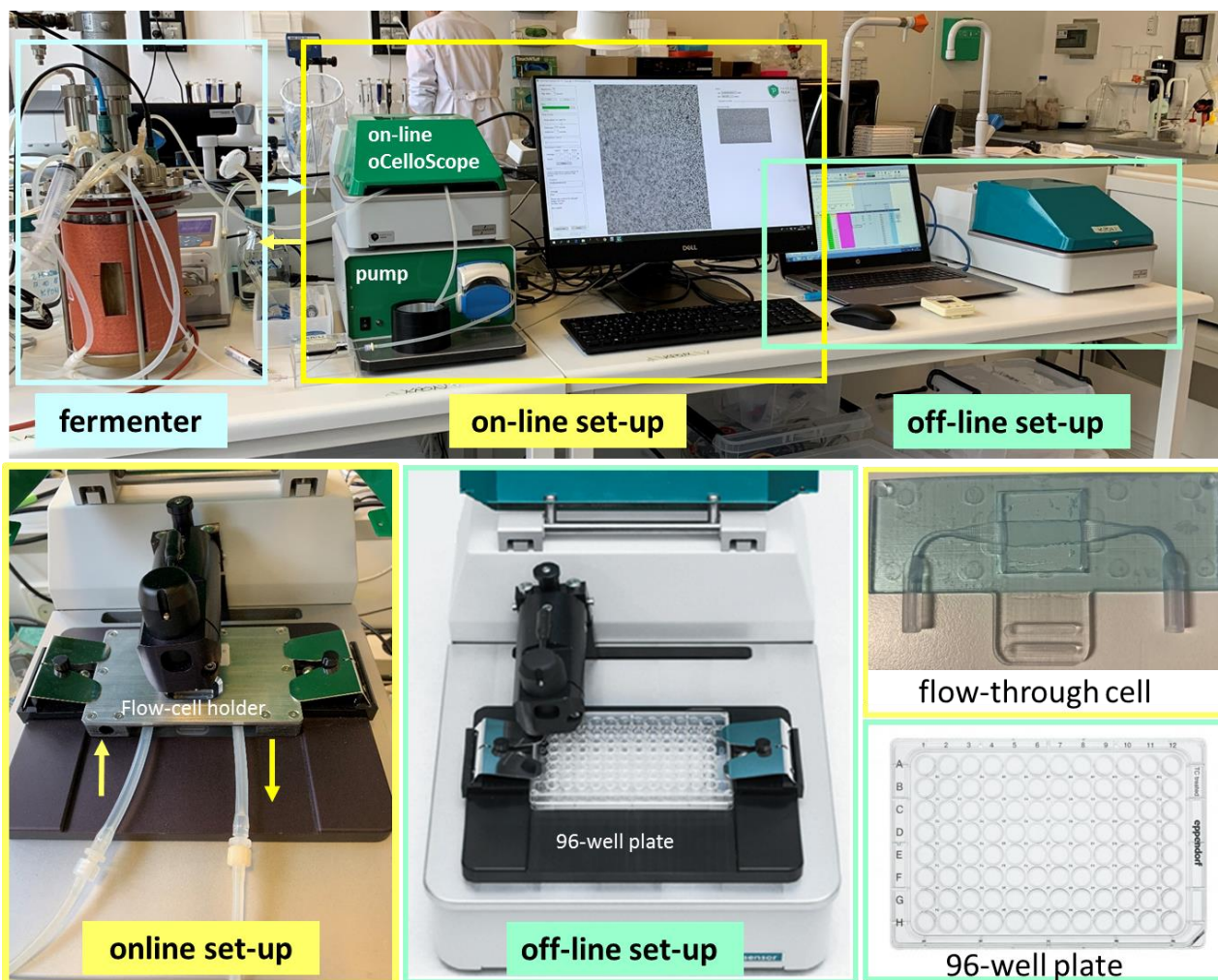


Figure 3.4: Fermentation set-up for imaging of the fermentation broth (on-line and off-line) including the fermenter, an oCelloScope unit connected on-line and operated in parallel off-line, respectively, and the different sampling devices for image acquisition by means of the oCelloScope (96 well-plate and flow-through-cell).

Both, on-line and off-line images were used for growth and morphological analysis. On-line images were collected automatically every 10 minutes via the flow-through-cell. Each 4th image was complemented with images acquired off-line on a 96 well-plate as described in 3.2.7 Off-line image acquisition. As the segmentation of the images acquired on-line could only be performed until a certain cell concentration (until 230 minutes of fermentation time), manual samples of defined dilution yielding a proper image quality for segmentation were collected off-line. Off-line imaging was performed in duplicates on both, an original (undiluted) fermentation sample and a sample of the diluted culture. In this way, growth could be studied on-line and off-line. On-line, segmentation could be performed until 230 minutes of fermentation time, but the data must be considered as critical due to too high cell concentrations even on the images acquired during the starting period of 230 minutes. However, based on sample dilution, images acquired off-line could be segmented all through the fermentation giving an insight into the morphological changes over the fermentation time. As cell-clusters in the fermentation broth might be affected by sample handling, sampling and dilution were conducted carefully and consistently following the same procedure.

In total, 57 images were acquired on-line (every 10 min, starting from 40 min of fermentation time) out of which the first 20 could be segmented and were considered for morphological investigations. All images were considered for growth detection.

15 images were collected off-line (every 40 min, starting from time zero) on an undiluted (natural) and diluted fermentation sample, respectively. The images collected on the undiluted samples were primarily used for growth analysis. Additionally, the first 6 images of the undiluted fermentation samples were considered for morphology dynamics. All 15 images collected on the diluted samples were used for morphology description.

[Table 3.1](#) summarizes the experimental differences between the images collected on-line and off-line, respectively.

Table3.1: Summary of the experimental and image parameters for image collected on-line and off-line during a 10 h yeast fermentation

	On-line images	Off-line images
Sampling device	flow-through-slide (prototype)	96 well-plate
Image size	Image distance: 45 μm No. of images: 100 Rows: 2 2.1 mm^2	Image distance: 4.9 μm No. of images: 50 Rows: 1 0.52 mm^2
Scanned Volume	9.3 nl	2.3 nl
Start after	40 min	0 min
Image collection rate	every 10 min	every 40 min
End after	600 min	600 min
No of images acquired	57 (0 – 56)	15 (1- 15)
No. of images segmented	20 (40 min – 230 min)	Undiluted: 6 (0 min -200 min) Diluted: 15 (0 min – 560 min)

3.3 Results and Discussion

3.3.1 The Fermentation Process

The progress of the yeast fed-batch fermentation process was described based on OD and continuous CO₂ measurements in the off-gas as well as by the yeast specific performance parameters of glucose (substrate) and the over-flow metabolic products ethanol, glycerol and acetic acid. While the concentration profiles of glucose, ethanol, glycerol and acetic acid are presented in Figure 3.5 A, the OD and CO₂ profile are presented in Figure 3.5 B. Additionally, the correlation between the typical yeast performance parameters glucose, ethanol, glycerol and acetic acid and the OD value is presented in Figure 3.5 C-F. The correlations of the different yeast performance parameters and the microbial growth were visualized with respect to the OD value. However, the correlations could possibly be based on the growth value obtained by the BCA, TA and SESA algorithm (image analysis). In this way, an image might be used to determine the biomass concentration and any of the performance parameters of interest simultaneously.

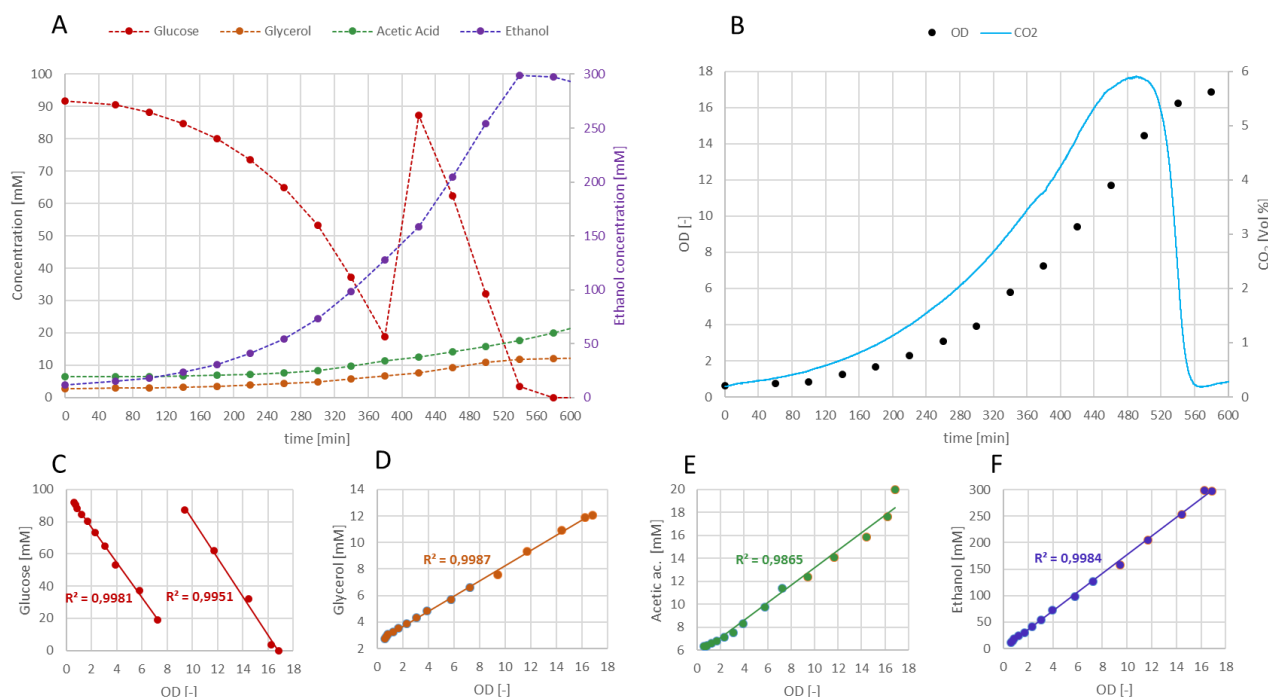


Figure 3.5: A) Concentration profiles in mM measured by HPLC for glucose, glycerol and acetic acid (left y-axis) and ethanol (right y-axis) as a function of the fermentation time in min. The standard deviation of HPLC measurements was < 2 %. B) Progression of yeast growth as indicated by the OD value (left y-axis) and CO₂ profile obtained from off-gas measurements (right y-axis). The standard deviation of the OD measurements was below 5 %. C – F) Correlation between glucose, glycerol, acetic acid, ethanol and the OD value. Note that the ‘jump’ in the glucose-OD correlation is the result of glucose addition after 380 min, correlating to the spike in glucose in Figure 3.5A.

Figure 3.5 A-B represent a typical progress of a yeast (fed-) batch fermentation. Glucose is consumed and transformed into biomass, CO₂ and ethanol, with acetic acid and glycerol as minor by-products. A final OD value of 17 is reached, representing a final cell dry mass concentration of approximately 8 g/l. Figure 3.5 B demonstrates the usefulness of CO₂ online data. Based on the CO₂ profile, the exponential growth phase can be followed and a small decrease in the slope can be used to detect the end of the exponential growth phase, characterized by glucose depletion. Hence, in this

experiment, the CO₂ data was used to decide the feeding moment of glucose. Moreover, [Figure 3.5 C-F](#) demonstrate that all analytes of interest inside the fermentation broth are highly correlated with the amount of biomass, or the OD value, respectively. Correlation coefficients of approximately 0.99 are yielded in all cases. The strong correlations between microbial growth, substrate consumption and metabolite production was referred to as the biological cage of covariance in chapter 2, section 2.2.1.2 Design of synthetic samples. The biological cage of covariance can be a challenge when those correlations need to be decoupled. However, when a reproducible process is in place, they may well be used for making indirect predictions, taking into account that the predictions won't account for process variation.

3.3.2 On-line fermentation images

The fermentation process under study was followed via imaging by means of the oCelloScope instrument every 10 minutes on-line and every 40 minutes off-line. As an example, 15 downsized images acquired on-line are presented in [Figure 3.6](#).

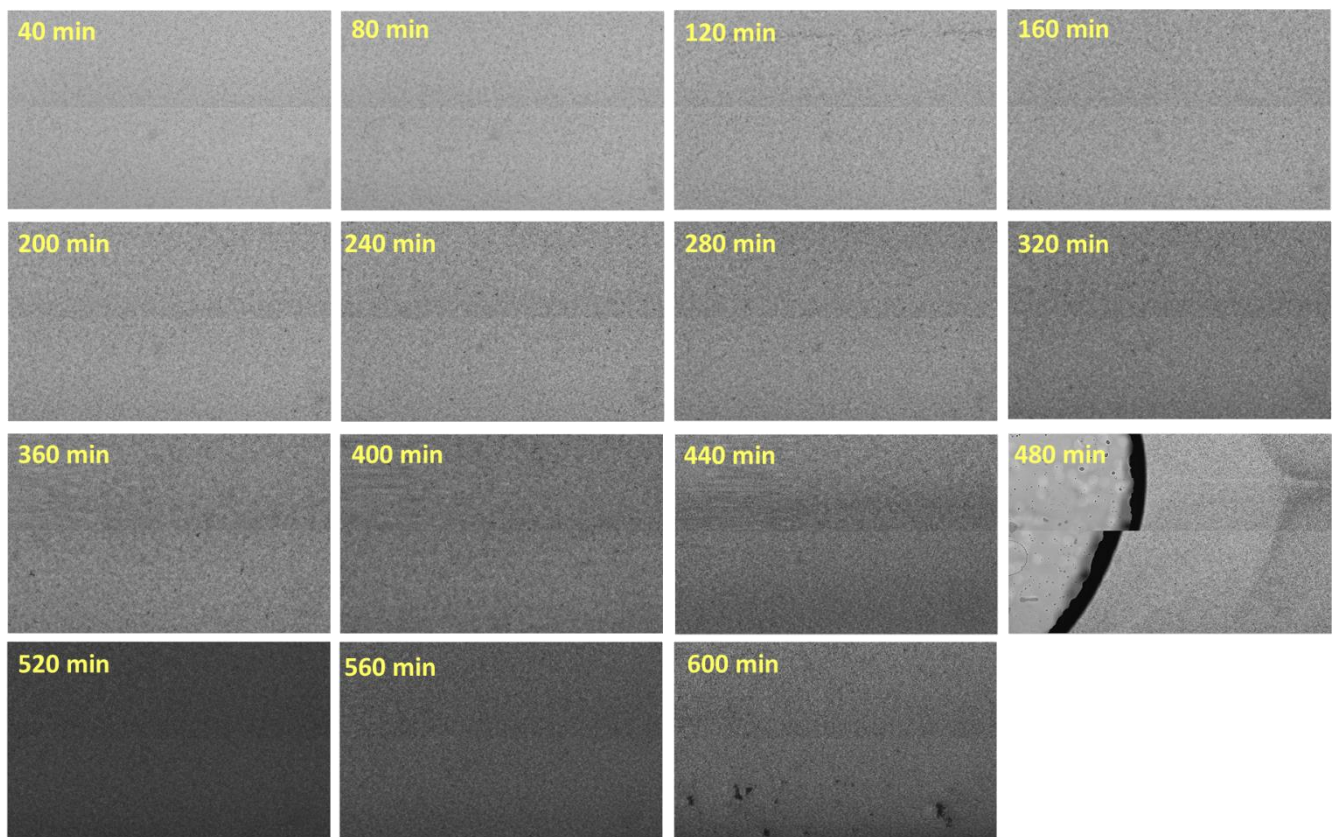


Figure 3.6: Downsized fermentation images collected on-line of 15 samples from start to end of the fermentation process. The best focus layer is shown for each image and the relative time point of image acquisition is noted.

[Figure 3.6](#) demonstrates the increase in absorbance caused by an increasing amount of cells over the fermentation time. The increasing amount of cells, progressively reducing the amount of light transmitted through the sample, essentially resulted in a grey scale of images. Moreover, the algorithmic joint of the first and second image row acquired resulted in a visible horizontal line in

the middle of each image. This can be regarded as an artefact resulting from the construction of the final image based on the 'patching' of the first and second image raw acquired. The image taken after 480 min clearly shows a disturbance. It can be related to a big air bubble stuck in the left end of the flow-through-cell. It is interesting to notice that the air bubble caused a visible streaming profile at the right-hand site of the image.

In order to see the images in more detail, a zoom into the images presented in [Figure 3.6](#) is shown in [Figure 3.7](#).

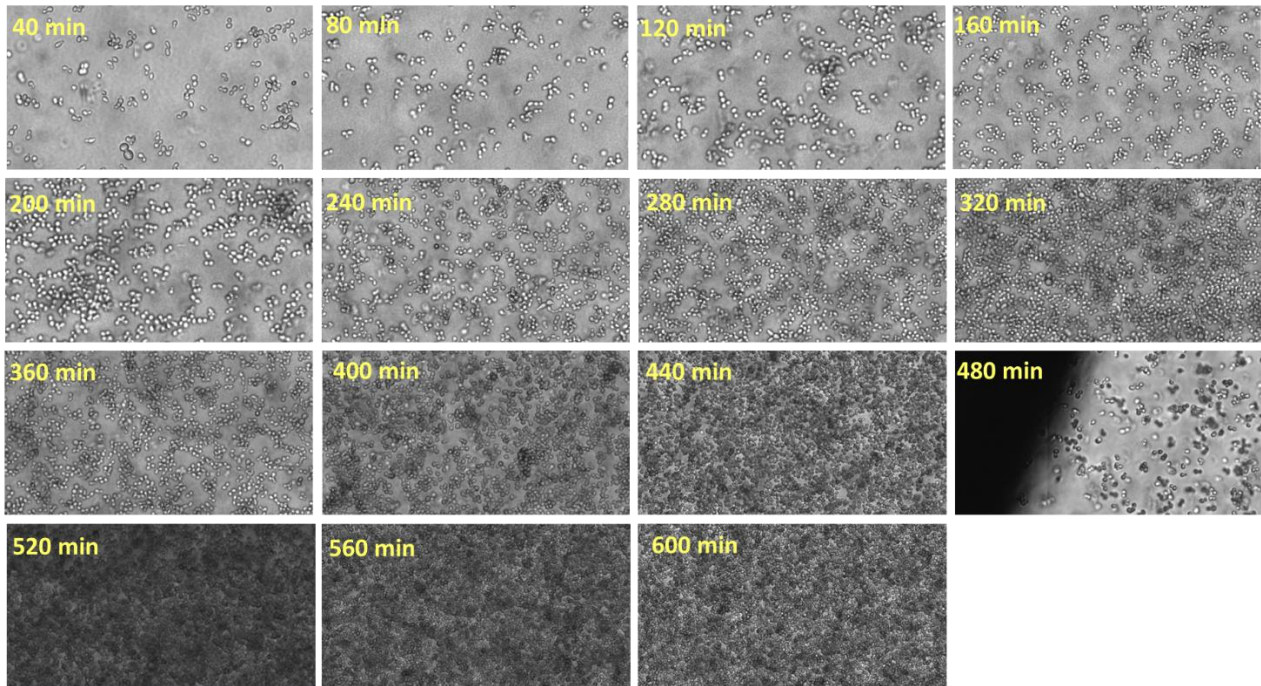


Figure 3.7: Zoom into the images acquired on-line presented in [Figure 3.6](#). The best focus layer is shown for each image and the relative time point of image acquisition is noted.

As long as the cell concentration is not too dense, yeast cells appear on the images as a bright spot surrounded by a darker border. Cell objects (single cells, budding cells, cell-clusters) on the images appeared as mostly separated from each other until 240 minutes of fermentation time. Afterwards, a distinction of different cell-objects became difficult or completely impossible even with the human eye. Finally, the yeast cells turned out as a mass progressively getting darker. In [Figure 3.7](#), the image acquired after 480 min represents the zoom into the respective image in [Figure 3.6](#). The dark shade dominating the left-hand third of the image represents the outer border of the air bubble appearing as a very dark, bold border in [Figure 3.6](#).

[Figure 3.8](#) demonstrates 6 segmented images within the time frame the segmentation algorithms were valid for the images acquired on-line. Due to the appearance of the yeast cells on the image as a bright spot surrounded by a darker border, a contrast and gradient based algorithm enables their detection. The algorithm used for segmentation of the on-line images underwent an area restriction.

If cells occupied more than 70 % of the image, the algorithm was programmed not to perform. Hence, segmentation of the on-line images could only be performed until 240 min of fermentation time. The segmented objects are colored randomly. One object appears in one color, and cells identified as one object appear in the same color.

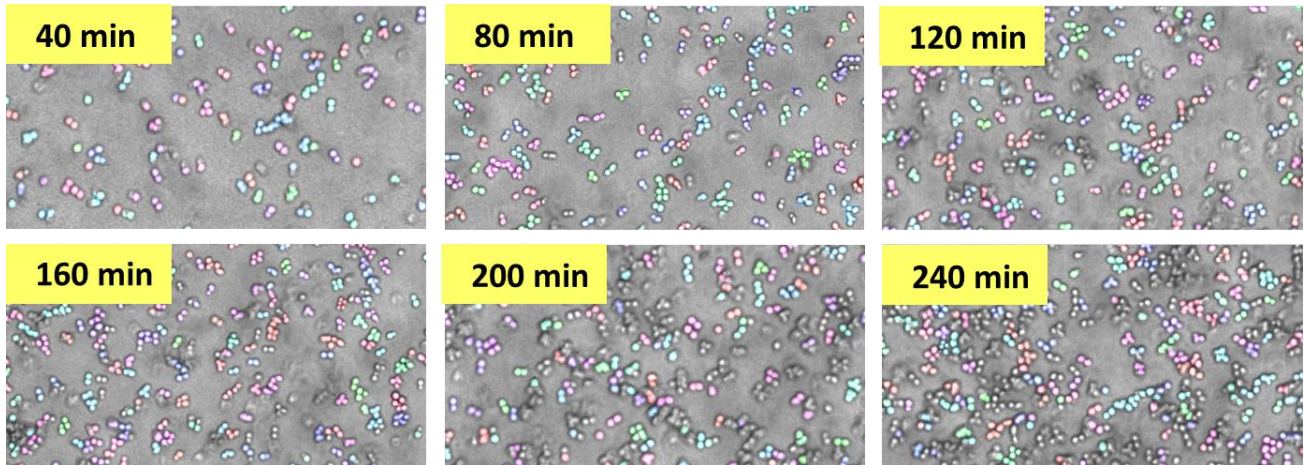


Figure 3.8: Segmentation of images acquired on-line. The relative time of acquisition is noted on each image. Images were segmented if cells occupied less than 70 % of the image area, which was the case until 240 min of fermentation time. Segmented objects are colored randomly, while each object is represented by one color.

As a first impression, the images and the performed segmentation appears fine. However, the image data collected on-line cannot be considered reliable with respect to the segmentation. This could be explained by too crowded images and out-of-focus cell-objects. The bottom of the prototype flow-cell can be considered as a foil, which tilted (deformed) more and more during the experiment. This affected the focus during the application, resulting in a lot of out-of-focus areas on the images. Cells out-of-focus cannot be segmented properly. Hence, shapes segmented did not represent the actual case, or in other words, cell-objects as a human eye would have identified them. In particular, the images acquired after approximately 120 minutes must be considered as too crowded. Too crowded images resulted in the following two issues. On the one hand, a too crowded image in- or out-of-focus produced a segmentation of large objects due to the presence of closely adjacent cells, which did not represent an actual cluster. On the other hand, too crowded images in-focus or out-of-focus produced a lot of small segmented objects showing one bright spot and were considered as single cells by the bright-spot algorithm, although they were actually part of a dense cell-cluster. These issues are demonstrated on a close zoom in [Figure 3.9](#), showing out-of-focus and adjacent objects with erroneous segmentation.

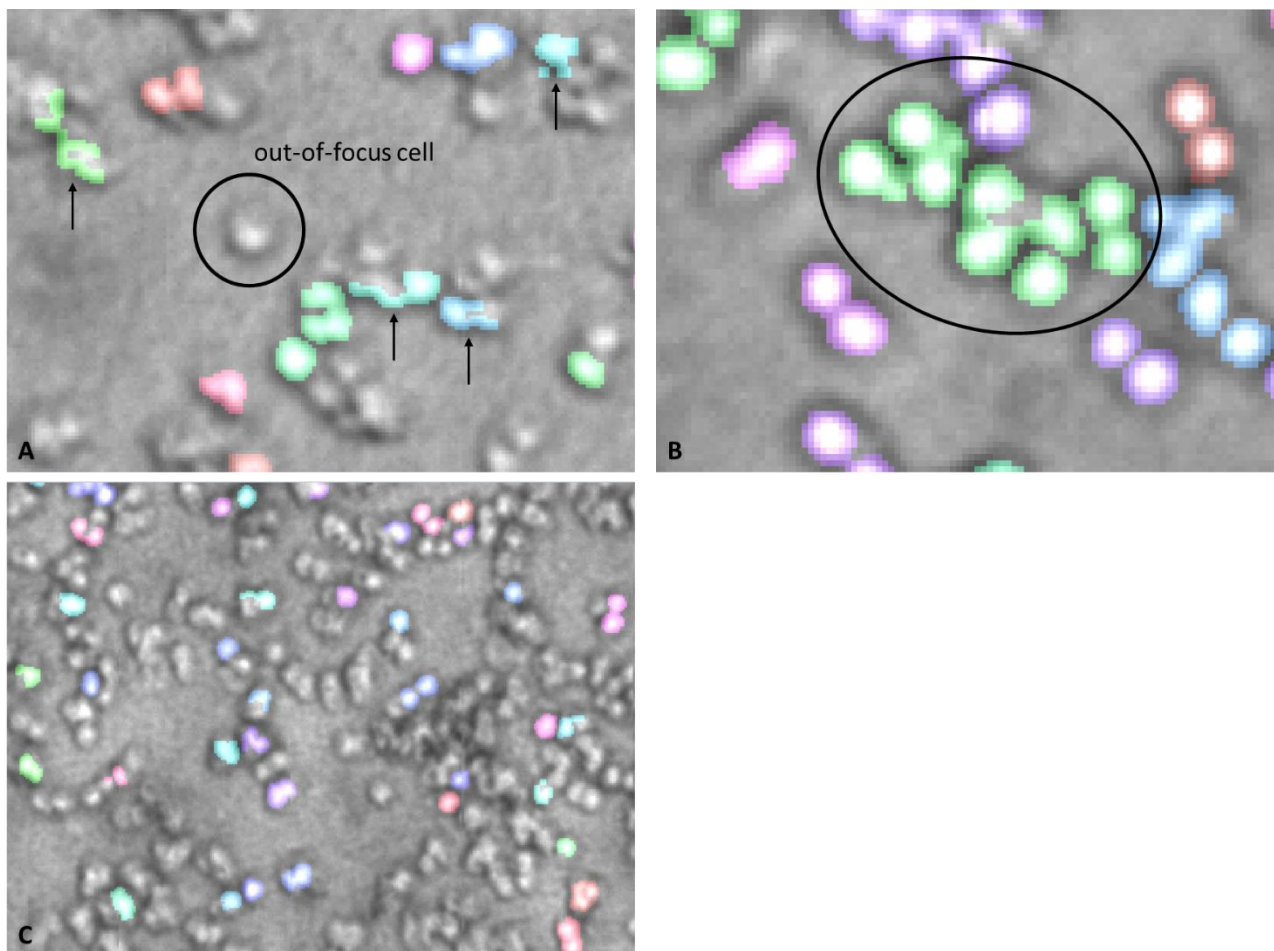


Figure 3.9: Demonstration of erroneous segmentation of images acquired on-line. A) An out-of-focus cell is marked with a circle. Segmented out-of-focus objects are marked with arrows. B) The orbited green cells are segmented as one object, due to adjacent border pixels. The green cluster must be considered as a segmentation artefact and not an actual cell cluster. C) A crowded, out-of-focus image resulting in a lot of single bright spot objects, not representing an actual full-size cell.

Finally, images collected on-line can be taken exemplarily as a first demonstration of an on-line imaging application. They could be considered for growth (in case of BAC and TA detection) but could not be taken into account for reliable segmentation (object identification) due to high cell concentrations and over time more and more dominating out-of-focus areas appearing on the images due to a bottom deformation of the prototype flow-through-cell.

3.3.3 Off-line fermentation images

Off-line images were acquired every 40 minutes in order to complement every 4th image acquired on-line. Off-line, images were acquired from original undiluted and diluted fermentation samples. The original fermentation samples could be used for off-line (automated) growth detection and segmentation analysis until a certain cell concentration or, respectively, until 200 minutes of the fermentation time. However, as the segmentation performed on the on-line images, the images taken off-line of the undiluted samples cannot be considered as reliable image material, due to too high cell concentrations on the images. The off-line images acquired on the undiluted fermentation

samples exemplarily shown on sample 1 out of 2 (a duplicate was acquired for each sample) for each time point are presented in [Figure 3.10](#).

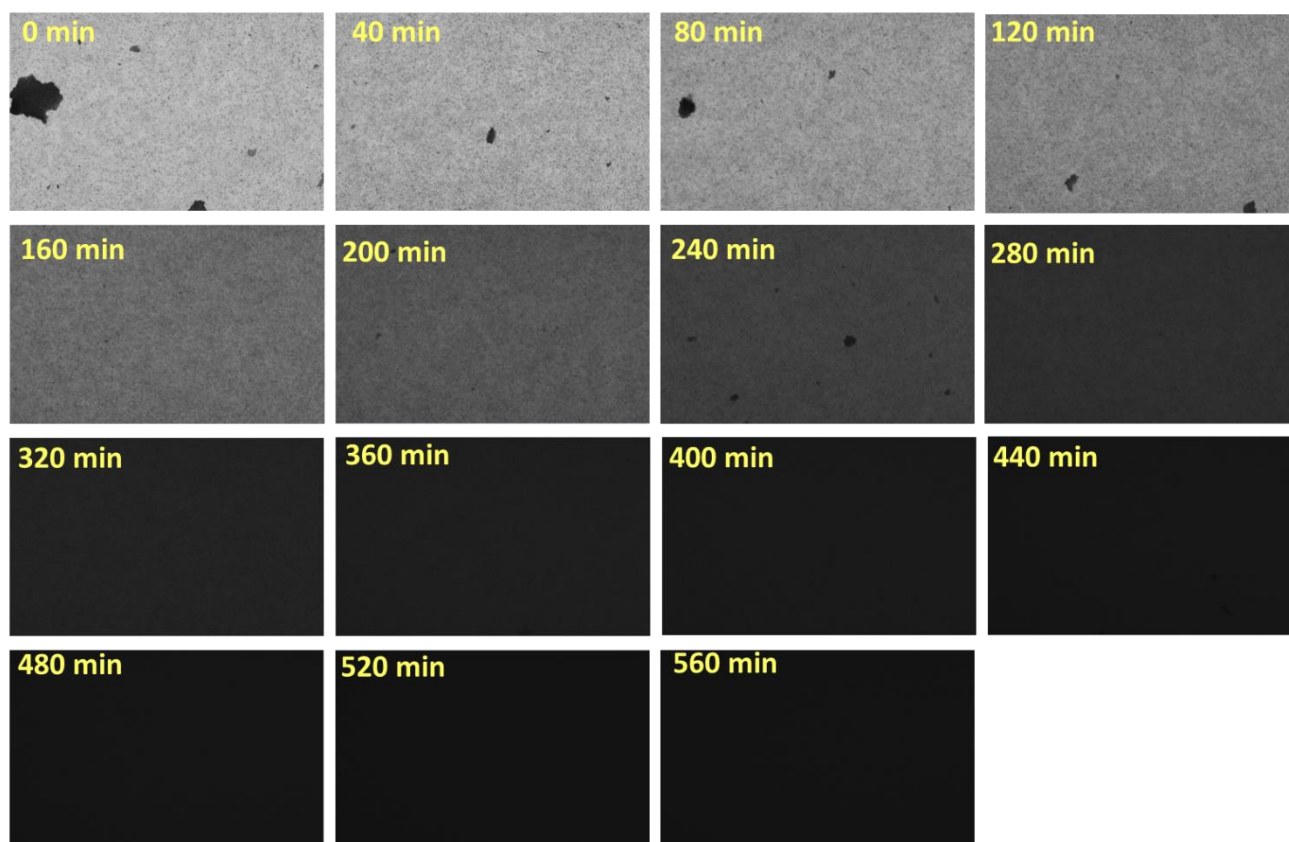


Figure 3.10: Downsized fermentation images collected off-line of all 15 samples taken during the fermentation process. The best focus layer is shown for each image and the relative time point of image acquisition is noted.

As the images acquired on-line, the images became progressively darker due to the increasing amount of cells. After 240 minutes, no change of texture is visible anymore to the human eye and after 240 minutes, the images appear as a solid dark background. Note that, the degree of absorption is generally higher as the sample layer inside a well (100 μ l sample volume) is thicker than inside the flow-through-cell. Hence, the images acquired off-line appear generally darker than the images acquired on-line. Some irregular dark spots can be observed in the first 4 images. A zoom into these dark spots led to the assumption that they were the result of very dense cell-clusters as exemplarily shown on a zoom into the dark spot at time point zero, presented in [Figure 3.11](#). [Figure 3.11](#) also reveals that the dark spot is not segmented as one object, but as some 'random', irregular objects inside the spot and at the spot borders.

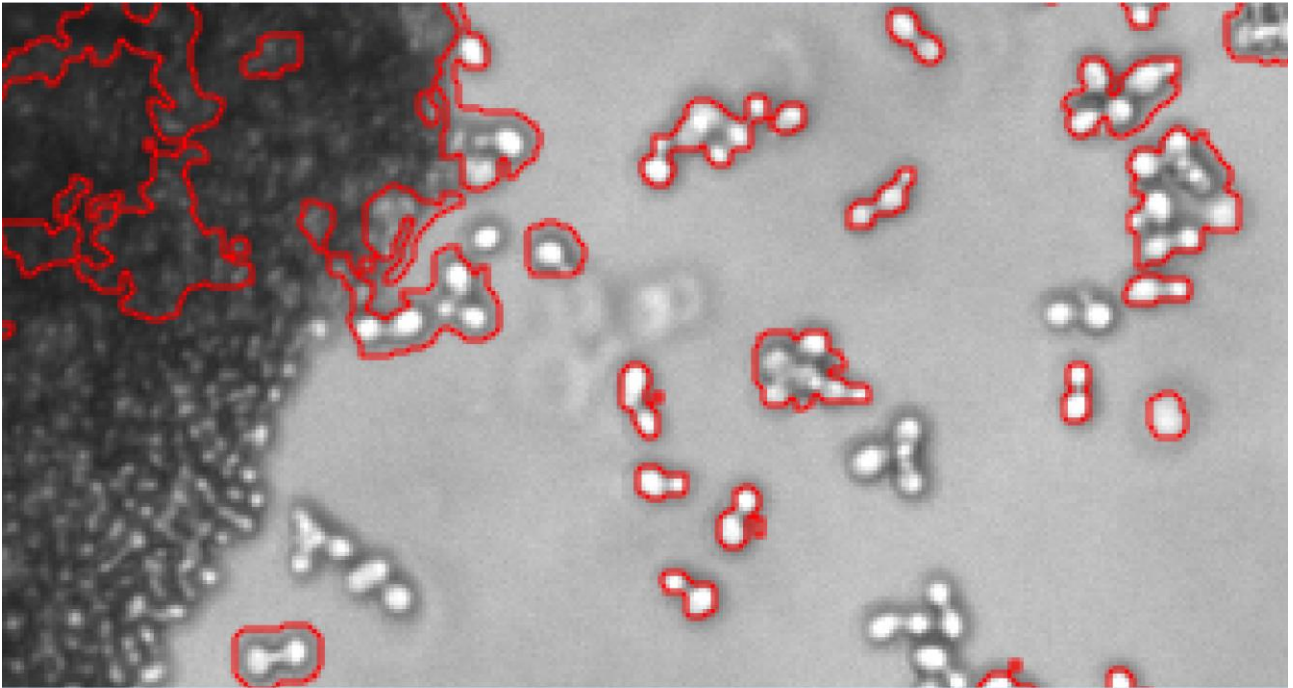


Figure 3.11: Zoom into the image acquired off-line at time point zero. The dark spot appearing at the left side of the image represents the border of the dark spot observed in Figure 3.10, at time 0. It seems to be a very dense cell-cluster.

A zoom into the off-line images acquired between 0 – 200 minutes on which segmentation could be performed as well as the respective segmented images are presented in Figure 3.12 A and B.

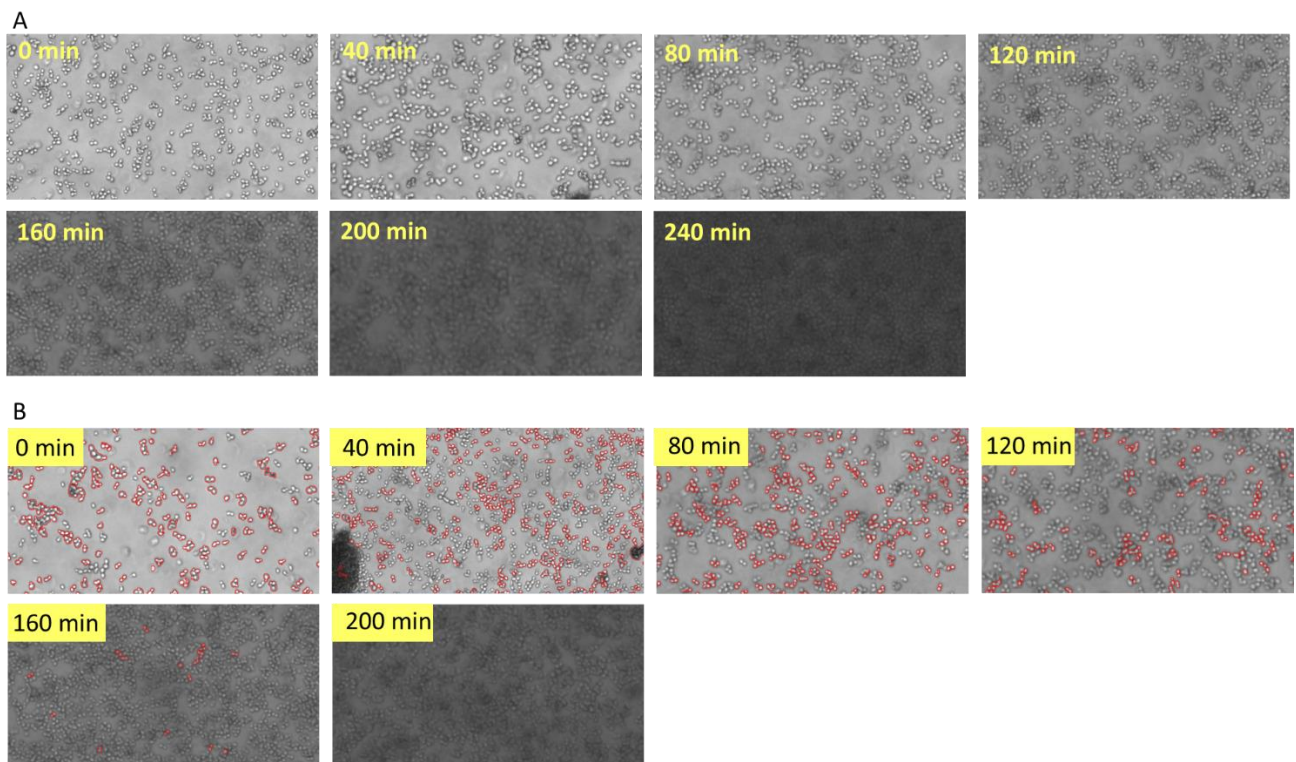


Figure 3.12: Zoom into the raw images (A) and segmented images (B) acquired off-line on which segmentation could be performed. Segmented objects are surrounded with a red line in Figure B). The best focus layer is shown for each image and the relative time point of image acquisition is noted.

The segmentation algorithm used on the off-line images segmented a maximal number of 5000 objects. As can be seen in [Figure 3.12](#), already the first image acquired off-line did not segment all the cells (objects), meaning that more than 5000 objects were present on the image. Generally, also the natural fermentation samples imaged off-line must be considered as too crowded for reliable segmentation. However, the focus was appropriate in place for all the images. Segmented objects are surrounded by a red line and no color assignment is part of this software version. An overview of the images acquired off-line on the diluted fermentation samples is given in [Figure 3.13](#).

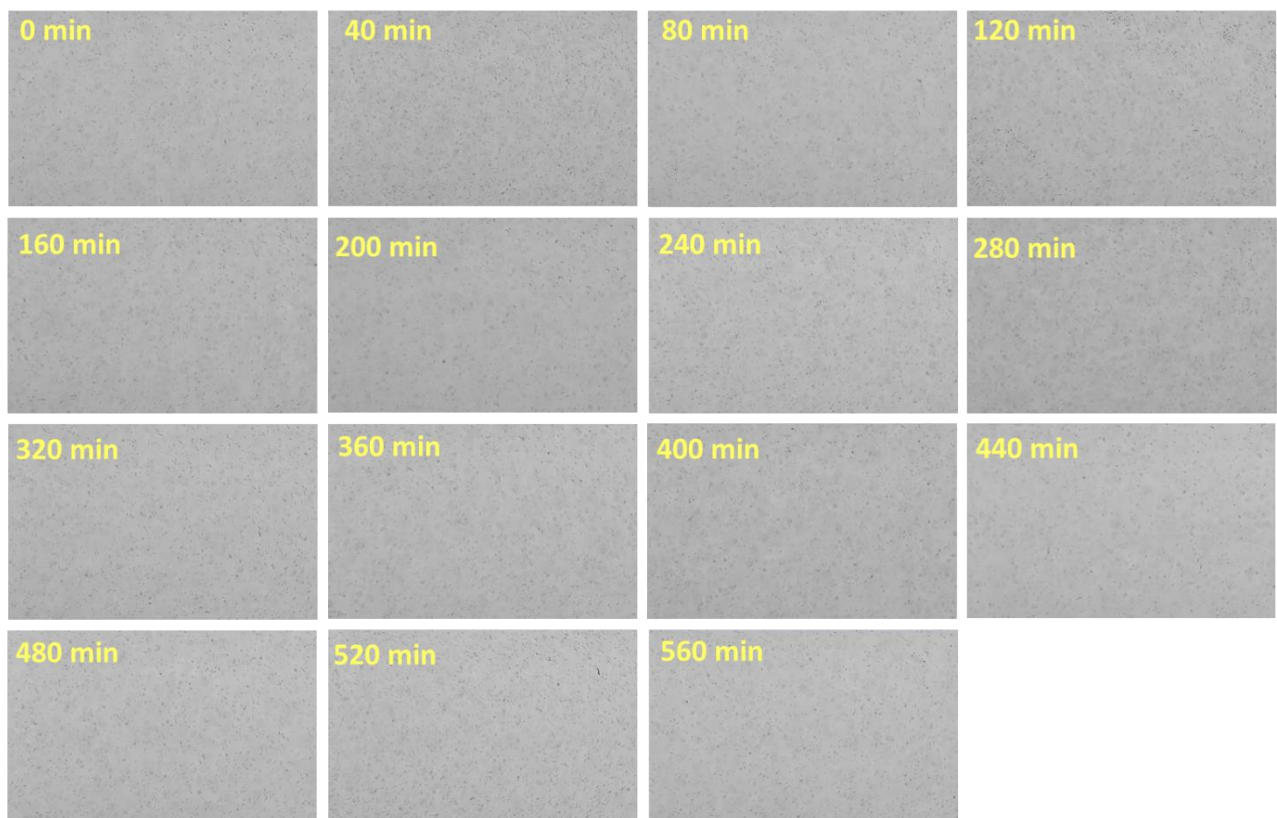


Figure 3.13: Downsized fermentation images collected off-line of all 15 samples taken during the fermentation process. All samples were diluted to an OD value of 0.1. The best focus layer is shown for each image and the relative time point of image acquisition is noted.

As all the samples possessed approximately the same cell concentration, [Figure 3.13](#) shows a uniform grey (absorbance) level for all images acquired off-line on diluted fermentation samples. No disturbances can be observed on the images. The image quality of the 15 samples can be regarded as equally good. Exemplarily, a zoom into the image acquired towards the middle and the end of the fermentation (after 240 and 520 min) on the diluted samples, as well as a segmented image section for these time points are given in [Figure 3.14](#).

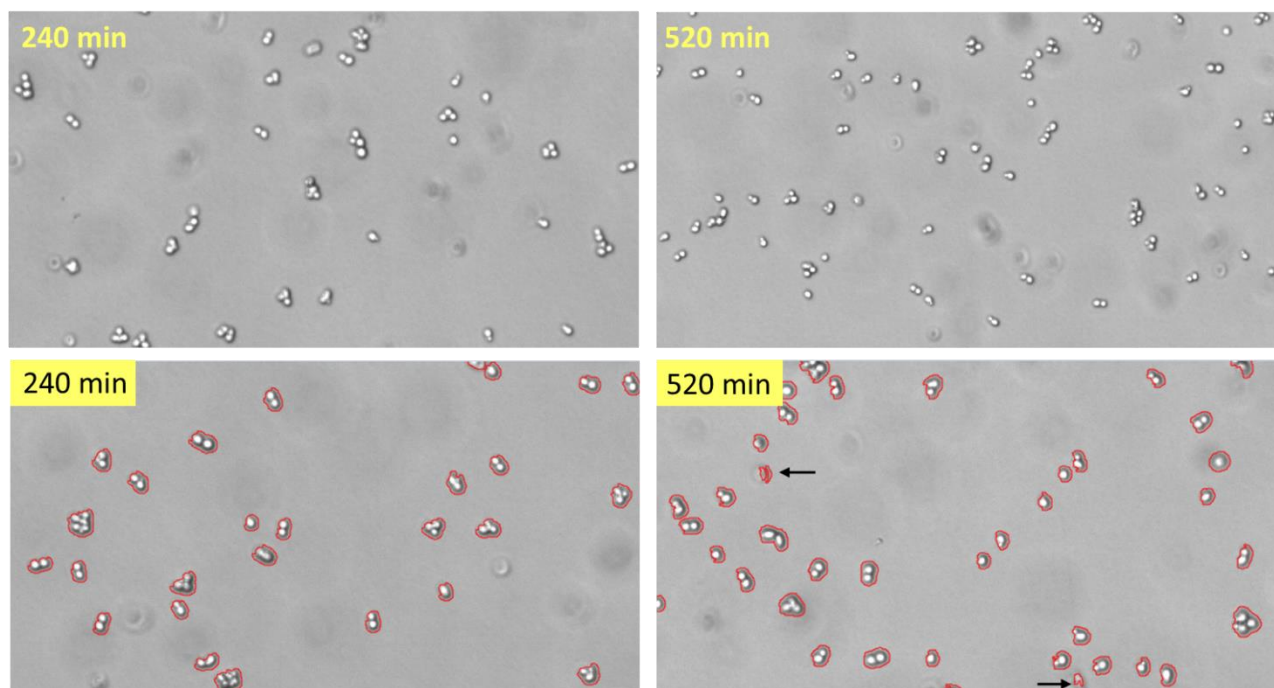


Figure 3.14: Zoom into the images acquired off-line of the diluted samples taken after 240 and 520 min (upper figures) and a segmented section of the respective images (lower figures). In the segmented image after 520 min, two arrows point to cells that are out-of-focus in the background and hence segmented deficiently.

According to [Figure 3.14](#), images were properly in focus and the cell concentration was appropriate to perform a reliable segmentation, as objects showed a good distance to each other (and were in focus). However, on any image, artefacts based on out-of-focus cells in the background could be observe to a minor extent. Two of those artefacts are pointed out by an arrow in the segmented image after 520 min in [Figure 3.14](#).

It is important to notice that quantitative segmentation parameters as the area, perimeter, circularity and the number of bright spots of an objects, chosen to describe the object (cell) morphology in this work, can only be evaluated if a link to the visual impression remains. The quantitative segmentation results presented in the following can only be explained and understood when compared to the actual image and to the segmentation performed. Hence, it was chosen to present a representative number of images in different sizes to create a general impression of the visual data. Numbers describing an object would remain numbers if the visual impression was unavailable.

3.3.4 Segmentation statistics

As the morphology dynamics with respect to the change in area, perimeter, circularity and the number of bright spots per objects will be presented as the average value per parameter per image, it is important to know on how many objects the average (mean) value was based. As the scanned volume of images acquired on-line and off-line (9.3 nl and 2.3 nl, respectively, [Table 3.1](#)) was different due to the different sampling-devices used (flow-through-cell vs. 96 well-plate), the number of objects present and segmented varied significantly between on-line and off-line images. Besides, in the diluted samples, the effect of sample dilution inherently reduced the number of cells compared

to the original fermentation sample. It is worth remembering that, segmentation could only be performed on the first 20 images acquired on-line and the first 6 images acquired off-line on the original, undiluted fermentation samples. In the case of diluted fermentation samples acquired off-line, all images could be used for segmentation analysis. For the respective segmented images, on-line and off-line, the number of objects identified per image for the three data sets 'on-line', 'off-line natural' and 'off-line diluted' is presented in Figure 3.15. Segmentation was limited by an area constraint, so that only images with less than 70 % cells (area wise) were segmented.

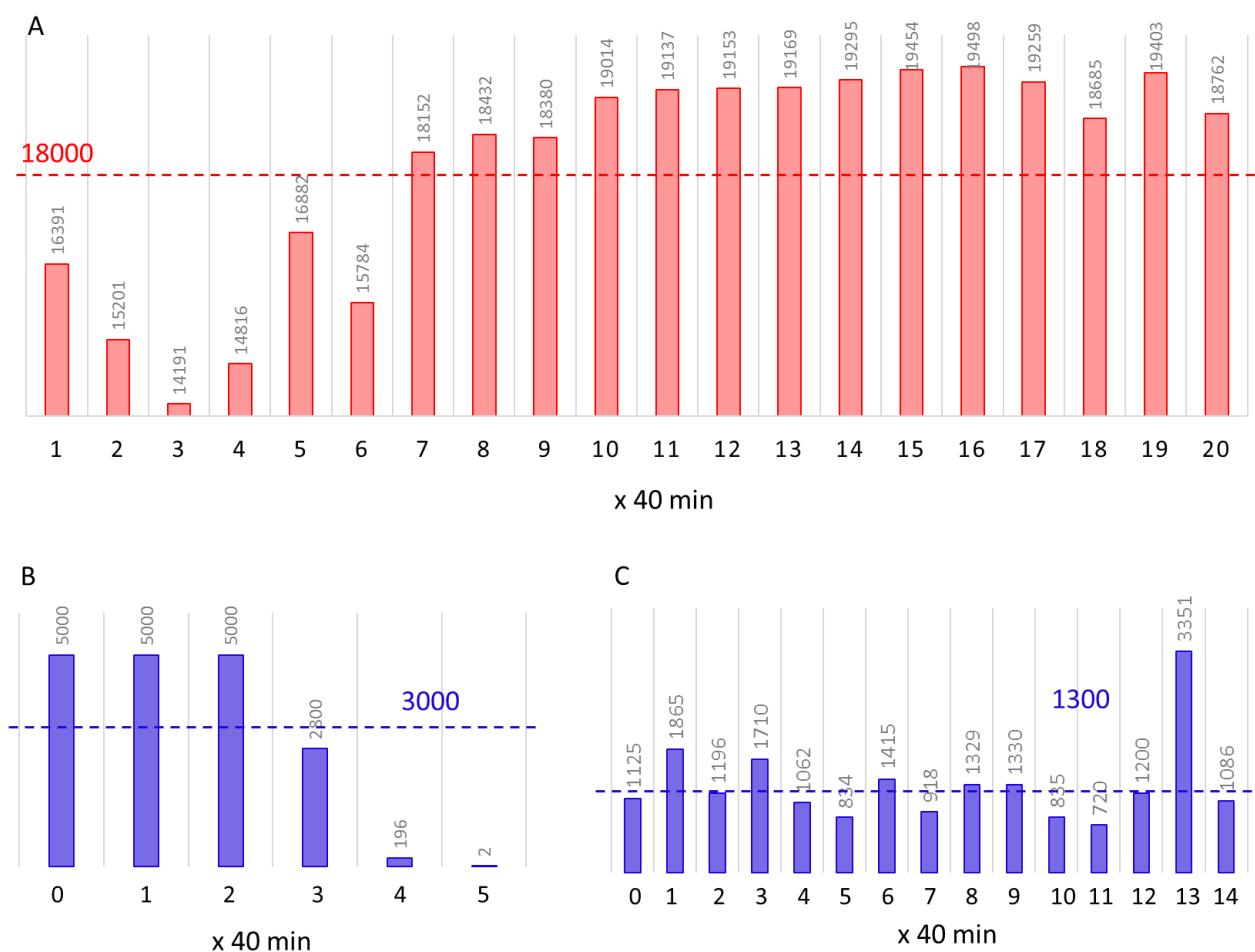


Figure 3.15: Number of objects segmented per image with respect to the on-line image data set (A) and the off-line image data set for (B) original undiluted and (C) diluted samples. The average number of objects rounded to the nearest 100 is indicated as a dashed line in each figure and the number is stated above the dashed line.

Moreover, the number of objects that could be segmented off-line was limited to 5000 objects per image. Note that, the number of objects identified does not represent the number of cells present on each image. Ideally, the number of objects identified represents the true number of cell-objects per image. Keeping in mind that a cell-object can be comprised by one or several cells, the algorithms as they are cannot be used for cell counting. However, they might be used for cell counting, in the case a feature is introduced that counts the total number of bright spots per image. As the cell concentration was from the beginning too high to yield a representative snapshot of the cell

morphology and the amount of cells, the number of objects identified did not include all the cell- objects present on the image, in case of undiluted samples (off-line and on-line). This could be observed in [Figure 3.8](#) and [3.12](#).

According to [Figure 3.15](#), 18000 objects were identified on average on each image acquired on-line ([Figure 3.15 A](#)). Around 14000 objects represented the minimum number of objects per image, and 19500 the maximum number. This was an order of magnitude higher than the number of objects identified off-line and is explained by the different scan volumes, or respectively, areas of the images acquired on-line and off-line. Off-line, on average 3000 objects were identified on each image, in case of the undiluted fermentation samples, while the maximum number of objects (5000) was reached on the first three images ([Figure 3.15 B](#)). Note that, in case of the undiluted off-line samples, on the last two images, only 196 and 2 objects were identified which cannot be considered as statistically representative in case of morphology analysis. In case of the diluted fermentation samples, 1300 objects were segmented on average ([Figure 3.15 C](#)). On the images of the diluted samples, no extreme outliers with respect to the number of objects identified per image were present. All images could be considered as statistically relevant. Although the amount of yeast cells was increasing over the fermentation time, the number of objects identified decreased with time, in case of the undiluted off-line samples ([Figure 3.15 B](#)). This did not represent the actual case in terms of fermentation progress. It demonstrates once more, that too crowded images cannot be considered for reliable segmentation. As over time, cells turned more and more into a grey mass on images acquired on the natural fermentation samples, the algorithms could not perform any longer due to a loss of image structure.

Another aspect regarding the number of objects is that the mean value per segmentation parameter might change dependent on how many objects were used for its calculation. This is expected to be in particular the case if the image quality is bad, in which case the results cannot be considered reliable. Bad image quality has so far been described as an image which is too crowded with objects and / or out-of-focus. [Figure 3.16](#) shows the mean value for the object descriptors chosen (area, perimeter, circularity, bright spots) as a function of time, dependent on the number of objects used for its calculation. The graphs shown in [Figure 3.16](#) are based on the on-line image data.

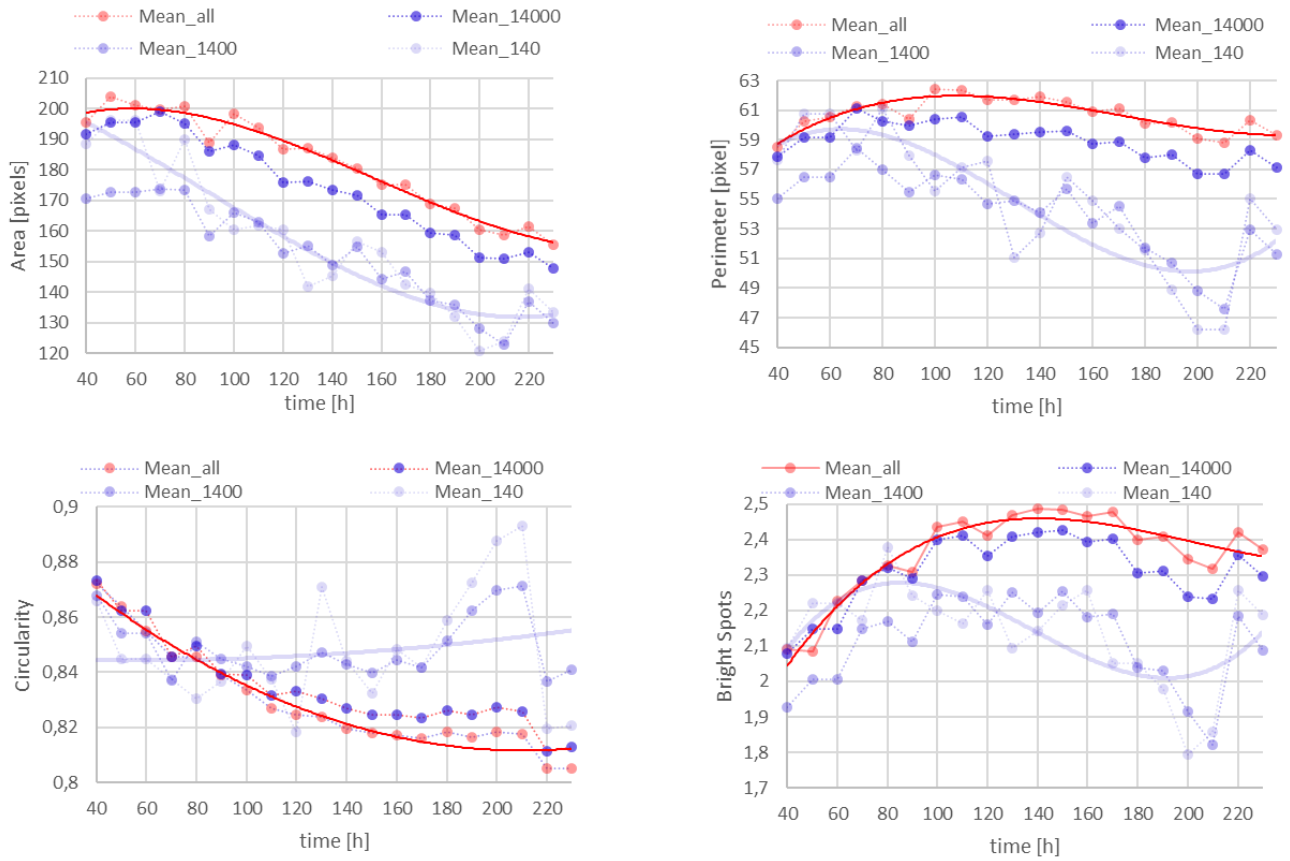


Figure 3.16: Mean values of the objects descriptors area, perimeter, circularity and bright spots as a function of time, based on the on-line image data segmented. The different mean values were calculated for a number of objects as indicated by the number in the legend. The red and light blue solid lines in each graph represent the trend for mean values based on all objects identified (Mean_all, red solid line) and the trend of the mean values based on 140 objects (Mean_140, light blue solid line), randomly chosen across all objects segmented per image.

Figure 3.16 was considered for statistical investigations. The trend of each parameter is discussed in detail in the section 3.3.6 Morphology dynamics based on image analysis. The parameter trend based on the mean values of all objects identified, as well as based on 14000, 1400 and 140 objects randomly selected among all objects was considered. 14000 objects were decided as this was approximately the minimum number of objects segmented in the on-line data-set. Moreover, Matlab operations can only be automated on matrices of equal size, so it is practical to reduce the number of objects *e.g.* to the minimal value found, without the loss of information. 1400 and 140 objects were chosen to reduce the number by two orders of magnitudes, whereas 1400 objects per image were approximately segmented on each image acquired on each diluted sample off-line. According to Figure 3.16, the parameter trend did change based on the number of objects used for calculating the mean. It can best be observed by comparing the solid red and light blue line in each graph. The lines show the trend line based on the mean value of all objects identified on the image (red solid line) and based on 140 objects randomly selected among all objects identified (light blue solid line). This supports again the previous conclusion that the segmentation could not be performed properly. Thus, all objects identified might give a worse representation of the actual image content than a

selection of 1400 objects, as outliers and / or deficiently segmented objects affect the mean value to a minor extent, assuming that a relevant number of objects was segmented properly in any case.

Finally, considering that image data is heavy in terms of disc space, the image size for on-line acquisition might well be reduced. Around 1000 objects properly segmented are considered to give a representative insight into morphology dynamics.

3.3.5 Microbial growth based on image analysis

Based on image analysis, microbial growth was determined by the three algorithms BCA, TA and SESA for on-line images and images acquired off-line on the natural fermentation samples. The images disturbed by an air bubble stuck in the flow cell (470 min – 490 min) were excluded from the growth analysis. The growth could visually be followed for on-line and off-line images in [Figure 3.6](#) and [Figure 3.10](#), respectively. Visually, growth was represented as a progressively darker getting image, due to an increasing amount of cells absorbing more and more light. While the TA and BCA algorithm quantify growth based on the same principle as optical density measurements, in other words, change in absorbance (pixel intensity), the SESA algorithm quantifies microbial biomass based on the image area covered by the objects identified. Thus, the SESA algorithm relies on a proper segmentation while the BCA and TA algorithm does not. [Figure 3.17](#) shows the growth data yielded for the three algorithms for both, on-line and off-line images in comparison to the OD curve. The TA and BCA values obtained are shown in [Figure 3.17 A](#), while the SESA values are shown in [Figure 3.17 B](#). The performance of the three algorithms was evaluated based on how well they were in agreement with the OD measurements. With respect to [Figure 3.17 A](#), it becomes clear that the TA and BCA algorithm represented the growth well up to 300 min of fermentation time. During this time, the OD value increased from 0.6 to 4 ([Figure 3.5 B](#)). Generally, on-line and off-line data show the same increasing trend until 300 minutes. After 300 minutes, both growth curves flatten out and became a horizontal line, whereas the OD data kept increasing. For both algorithms, the off-line data resulted in generally higher values than the on-line data. This is in agreement with an overall higher absorbance in case of the off-line image data, which, in turn, can be traced back to a higher liquid layer in the 96-well plate compared to the liquid layer in the flow-through cell. Accordingly, the off-line images were found to be generally darker than the on-line images ([Figure 3.6](#) and [3.10](#)).

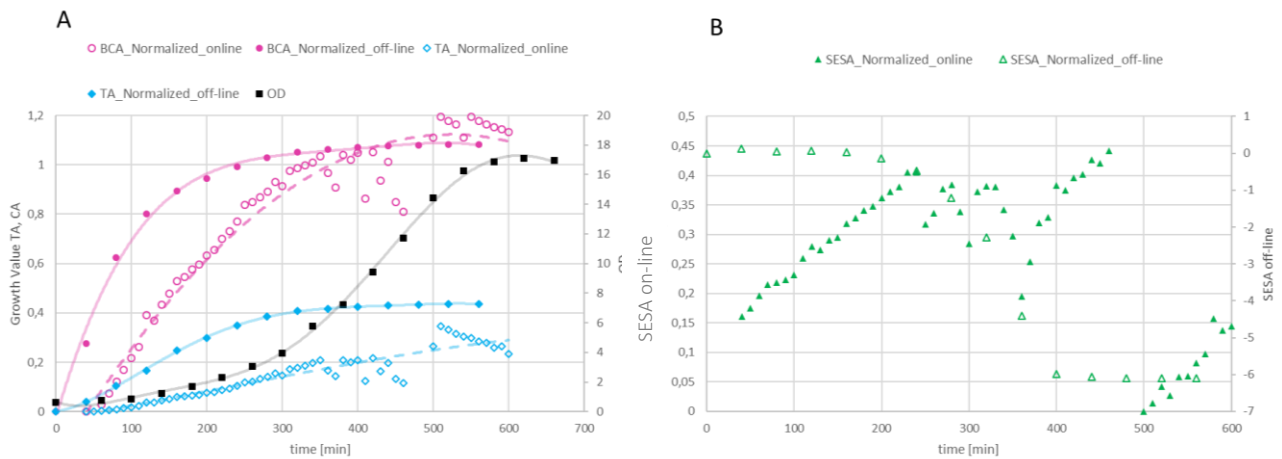


Figure 3.17: A) Microbial growth over the fermentation time based on image analysis applying the TA and BCA algorithm to the on-line and off-line images acquired from the natural fermentation samples. TA and BCA values are presented by the left-hand side y-axis, while OD data is shown for comparative reasons on the right-hand side y-axis. B) Microbial growth over the fermentation time based on image analysis applying the SEAS algorithm to the on-line and off-line images acquired from the natural fermentation samples. The left y-axis represents the SEAS values obtained on-line, the right y-axis shows the SEAS values obtained off-line.

Moreover, it is demonstrated that the BCA algorithm definitely showed increased sensitivity over TA and OD measurements, represented as an initial slope that was more than 7 times higher compared to the TA and OD increase.

With respect to the performance of the SEAS algorithms (Figure 3.17 B), neither on-line nor off-line images could be considered as reliable image material for growth detection within this study. The image quality did not allow proper segmentation as discussed before. However, with respect to the images acquired on-line until approximately 200 minutes, representing an increase in OD from 0.6 to 2, the on-line data showed an increase which is in agreement with the OD data. Hence, the algorithms could still measure an increase in the total object area covering the image. Contrarily, the off-line SEAS value stayed constant at approximately 0.45 until 200 minutes and decreased after 200 minutes. This can be explained by an image that was too crowded from the beginning for proper segmentation. The software used off-line was restricted to identify maximal 5000 objects. From the beginning, more than 5000 objects were present, as it could be observed on the images that not all objects were segmented (Figure 3.12 B). Hence, an increase in cells (and hence, total cell area) could not be caught as the upper detection limit was exceeded, resulting in a constant number of 5000 objects identified. The decrease in SEAS value for the off-line images after 200 min simply represents the decrease in image quality, allowing for less objects to be identified out of the cell mass.

It can be concluded that, for the early stages of growth, up to an OD value of 4 within this work using a 96-well plate and a sample volume of 100 μ l, and respectively a flow-through-cell with a height of 800 μ m, both BCA and TA algorithm could be considered as a reliable method for growth quantification. The BCA algorithm clearly showed increased sensitivity over the TA algorithm and OD measurements, respectively, resulting in a significantly higher slope compared to the TA and OD increase. However, both algorithms were not able to detect the ongoing increase in biomass after 300 minutes when OD values higher than 4 were reached. The upper limit of biomass detection might

be a bit increased by lowering the depth of the liquid layer of the sample for image acquisition. In case of the flow-through-cell, a device of 800 μm height was chosen (another type of 300 μm was available) in order to prevent the clogging of the flow-through cell when reaching higher cell-concentration during the fermentation. No issues in terms of clogging or interference with the flow were observed during the fermentation, and the height of the flow-through cell chosen for application might well be reduced. In the future, the flow-through-cell will also be available in a height of 140 μm . The height and hence, the absorbance of the sample might be reduced by a factor of 5.7 which in turn might allow for a detection of cell concentrations that are 5.7 times higher than in the current study. Thus, by reducing the height of the flow-through cell, OD values up to 22 might be detected by the BCA and TA algorithm and hence, yeast growth during the entire batch fermentation as in this study could have been followed. Considering the 96 well plate as a sampling device, the volume might be reduced to 50 μl , assuring that the entire bottom of a well is covered with sample liquid. That might have allowed to extend the detection of yeast growth to an OD value around 8.

The SESA algorithm could not be properly applied within this work, as the initial cell concentration (an OD of 0.6) was already too high to allow for proper segmentation on the images acquired.

The amount of biomass is strongly correlated to the concentration level (or, respectively, consumption / production rate) of substrate and products, as shown in [Figure 3.5 C-F](#). Within this study, this inherent correlation was not found to be particularly useful as the detection of growth by image analysis was limited to low biomass concentrations. However, considering that these limitations might be circumvented by the sampling device applied, an image might be used to yield results with respect to biomass detection and, via correlations, also an idea of the substrate/ product level might be gained when integrating the respective mechanistic models.

3.3.6 Morphology dynamics based on image analysis

All objects identified on each image were quantitatively described by several parameters linked to each object. Within this study, the objects were described by means of the area, perimeter, circularity and the number of bright spots. These descriptors were chosen in order to yield a representative and complementary data set describing the cell morphology under study. The following morphology dynamics might be expected. During the cultivation, the number of bright spots would increase, essentially from one bright spot representing a single cell, to two and more bright spots representing a reproducing cell or a cell cluster. More than two bright spots per object might describe a very high growth rate, where a bud starts to grow itself prior to its separation from the mother cell. Hence, the exponential growth phase might show mainly a lot of budding cells and small cell-clusters, while the lag phase or a phase of maintenance might be substantially described by a lot of single cells. An increasing number of bright spots per object is expected to come along with an increase in area and perimeter and respectively a decrease in circularity. The circularity is expected to decrease with an increasing number of bright spots due to the objects becoming more and more branched. A decreasing number of bright spots is expected to show the opposite trend. In order to describe the

morphology dynamics, the mean value for each parameter was calculated and plotted as a function of the fermentation time. [Figure 3.18](#) shows the morphology dynamics for both, at- and off-line image data. Note that, the off-line images were acquired as duplicates based on two different wells, as indicated in the legends. Dotted, dashed and solid lines represent the average data trend. The trend lines were approximated by polynomials. It is important to realize that these results represent the segmentation and in case of a bad image quality do not represent the actual object morphology. Keeping this in mind, only the images acquired off-line on diluted samples can be considered as reliable image material. The images acquired on-line and off-line on the natural fermentation samples must be considered as a bad source material for segmentation, as the images were out-of-focus and too crowded for representative segmentation.

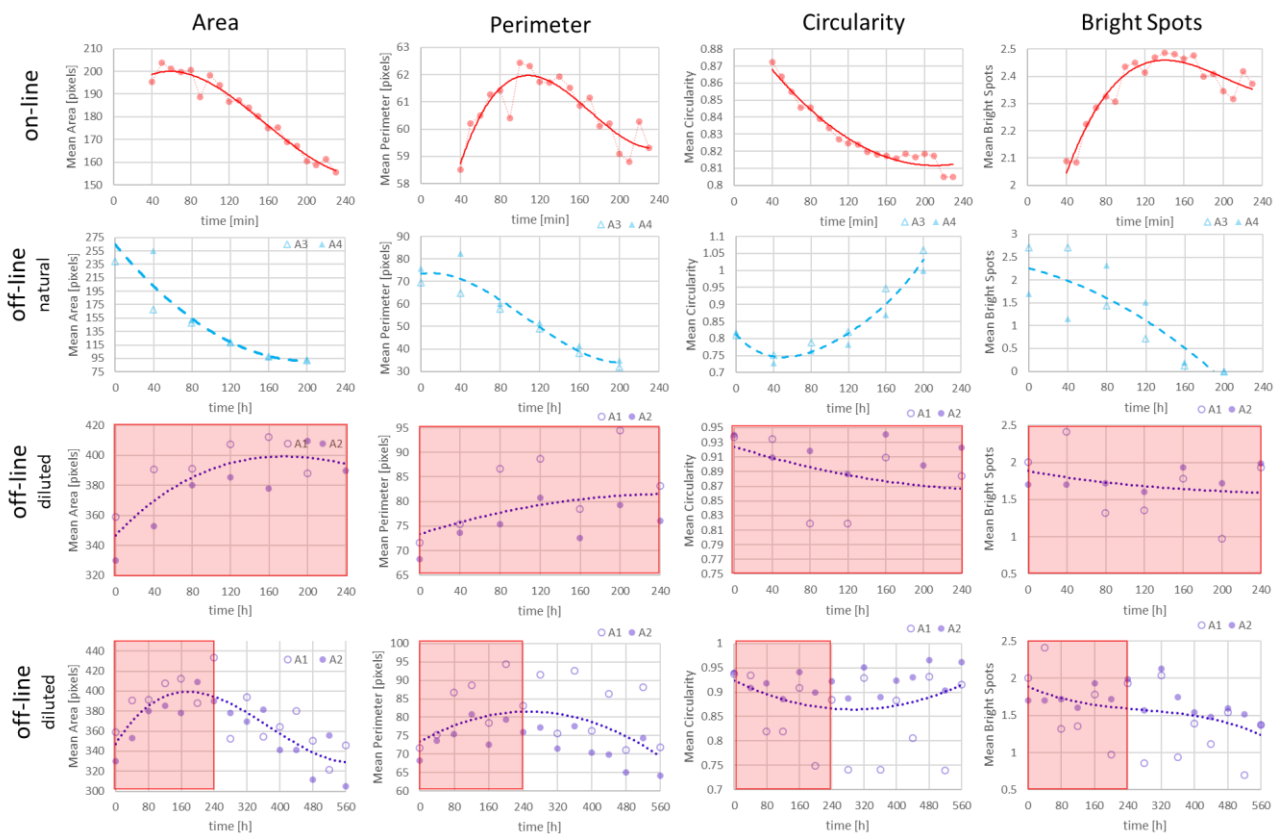


Figure 3.18: Morphology dynamics over the fermentation time for on-line and off-line image data with respect to the descriptors area, perimeter, circularity and number of bright spots. Each parameter is presented as the time trend of the mean value. Row three represents a zoom into the full-time data shown in row 4, as marked in red, in order to have a comparative insight in the parameter trends within the first 240 min, obtained on the three different data sets 'on-line', off-line (natural)' and 'off-line (diluted)'. The trend lines, represented as continuous, solid, dotted and dashed lines, were approximated by polynomials.

Generally, it can be observed that on-line and off-line image data resulted in contradicting trends. This could be expected due to bad image quality in terms of segmentation. The first 20 images acquired on-line were segmented until 230 min of fermentation time. According to [Figure 3.18](#), the following segmentation trends were obtained. Over the time during which segmentation could be performed, the area and circularity showed a continuously decreasing trend. The perimeter and the

number of bright spots showed the shape of a parabola, with a maximum observed at around 120 min. The mean object area decreased ca. 22 % (from approximately 200 to 155 pixels), the mean circularity decreased ca. 7 % (from 0.87 to 0.8). The perimeter varied ca. 5 % (maximal around 4 pixels), the number of bright spots varied ca. 20 % (between 2 and 2.4). Two conclusions can be drawn. Firstly, the changes in all morphology descriptors were minor. Secondly, the observed morphology trends were artefacts produced by deficient image quality resulting in erroneous segmentation. In a good data set, the descriptors chosen are expected to complement each other when describing cell morphology, in the way described in the beginning of this section. This was not the case here, *i.g.* the data is contradicting by itself. According to the segmentation data in [Figure 3.18](#) (on-line trends), the mean area and circularity decreased while the perimeter and number of bright spots increased. This cannot be true considering the natural shape of yeast cells, appearing as round objects and round objects 'stuck' together. Hence the trends observed must be considered as a description of what the algorithms segmented due to bad image quality, but not the actual yeast cell-morphology.

In particular the out-of-focus issue resulted in a segmentation of yeast cells which did not reflect the natural shape of yeast cells. The decreasing trend in area over the fermentation time is explained by objects with 'open' segmentation borders ([Figure 3.9 A](#)), thus resulting in a large perimeter but essentially no area. Hence, providing appropriate image quality, *i.e.*, images on which cell-objects appear separated from each other and in-focus, is crucial for yielding reliable segmentation results. The challenge of out-of-focus cells and high cell concentration was addressed by ParticleTech Aps, as described in 3.4 Conclusions and future perspectives.

The morphological trends observed based on the images acquired off-line on the natural fermentation samples showed data consistency, as it would be expected in nature. A decreasing number of bright spots came along with a decrease in object area and perimeter and an increase in circularity. In nature, this can be explained by a transformation from cells occurring rather in clusters to cells occurring rather as a single entities. However, the opposite of this trend is expected at the beginning of the fermentation process. Cells transitioning from the lag phase to the exponential growth phase are expected to 'transform' from single cells to (rapidly growing) budding cells, producing a lot of budding cells and eventually cell-clusters. Hence, also in the off-line data set yielded from natural fermentation samples, the trends found describe the deficient image quality resulting in erroneous segmentation and not the actual cell morphology. However, the data is more consistent in itself, as the images were properly in focus. This reveals that crowded images in-focus produce less artefacts with respect to the natural morphology trends expected, than out-of-focus images.

The off-line images acquired on the diluted samples were considered as good image quality with respect to the segmentation. The images were in focus and objects appeared in good distance to each other. The time trends of mean object area, perimeter and circularity occurred as expected. An

increase in area came along with an increase in perimeter and a decrease in circularity. Nevertheless, the number of bright spots was decreasing according to the trend line. However, the duplicates did not produce the most consistent data, and approximately every second data point yielded from the duplicate (indicated as A1) seems to contradict the actual trend.

It can be concluded that on-line and off-line image data obtained from the undiluted fermentation samples yielded deficient segmentation due to bad image quality. In this case, the trends observed represent what was segmented by the algorithms and not what was actually present on the image. Segmentation was deficient due to out-of-focus and too crowded images. In case of the diluted fermentation samples, off-line image data could be used for segmentation and hence, a first insight into the morphology dynamics of yeast cells during a (fed-) batch fermentation could be gained. For the mean values of area, perimeter and circularity over the fermentation time, the morphology trends described a parabola. The area and perimeter of the objects were first increasing on average, followed by a decrease to approximately the starting value. The opposite trend was observed for the circularity. The number of bright spots was found to be continuously decreasing. However, it is assumed that the bright spot trend obtained was due to bad replication (deviations between the first and second image sample) and a parabola shape as for area and perimeter might be anticipated. The curve maximum in area and perimeter, the curve minimum in circularity, respectively, occurred approximately after 220 min of the fermentation time. The initial 220 min of fermentation represented the early stages of the exponential growth phase, and the growth rate can be considered as still rather low (Figure 3.5 B), while glucose concentrations, in contrast, are still rather high (90 mM - 75 mM, Figure 3.5 A). Interestingly, these findings suggest the opposite of the morphology progression assumed at the beginning of the chapter. Cells tend to 'cluster' at rather low growth rates, and not as expected, at high growth rates. Morphology dynamics described as the mean value per time point (image) changed marginally. More experimental and reliable image data must be collected and analyzed in order to confirm the first findings. However, yeast cells seem to undergo morphological changes during growth. If these changes can be captured reliably and correlated to the crucial process parameters of interest, monitoring and control of fermentation processes based on imaging and image analysis will become an exciting novel strategy.

3.3.7 Performance of the bright spot feature

The bright spot feature was developed within this work aiming at the automatic distinctions between single cells, budding cells and cell-clusters of different sizes on the images. Its performance was evaluated in Figure 3.19. In Figure 3.19, all objects segmented on the images acquired off-line on the diluted fermentations samples, exemplarily shown after 0, 240 and 560 min, were automatically assigned to a group, defined by the number of bright spots. Each group (objects of 0, 1, 2, 3, 4, more than 5 bright spots) was manually checked for false objects, in other words, objects that a human eye would have assigned to a different group, or considered to be segmented erroneously, were identified and removed from the data set.

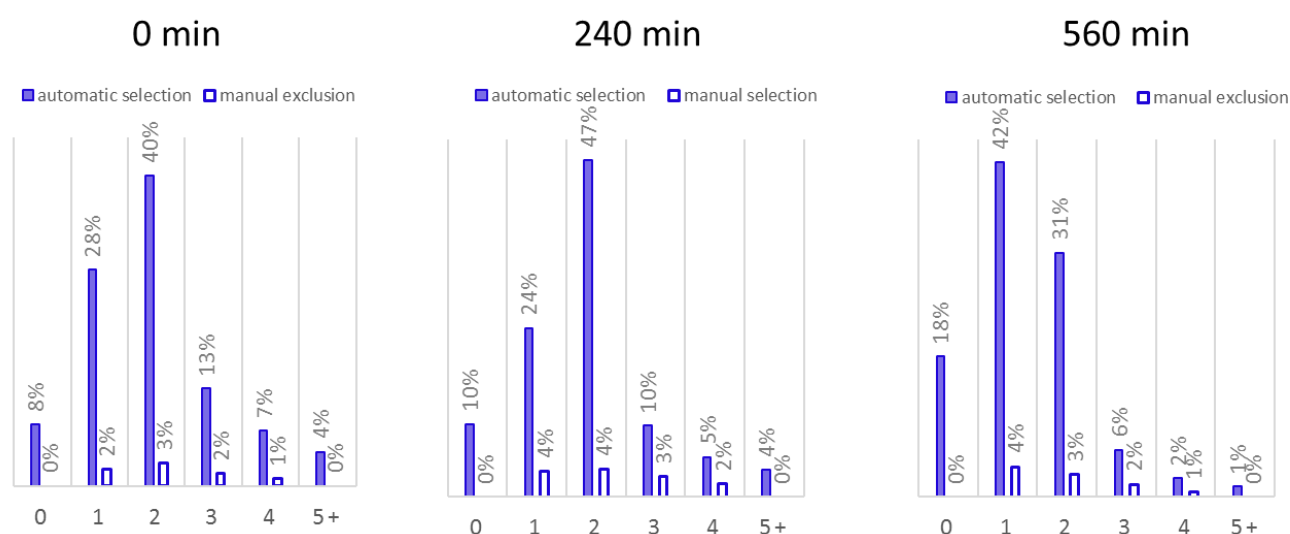


Figure 3.19: Evaluation of the bright spot feature exemplarily shown on the images acquired off-line on the diluted samples (replicate 1) after 0, 240 and 560 min of fermentation time. Based on the bright spot feature, every group of objects belonging to 0, 1, 2, 3, 4 and more than 5 bright spots was checked manually for false objects and objects erroneously assigned to that group were excluded manually.

Objects assigned to each group based on the bright spot algorithm can be found in Figures S3.1 – S3.6 in the supplementary material. Objects manually excluded from each group can be found in Figures S3.7-S3.10.

In general, segmented objects with zero bright spots can be considered as out-of-focus cells in the background or non-cell objects segmented due to ‘shadings’ on the image. The bright spot feature can be used for the automatic exclusion of objects segmented as a ‘zero bright spot object.’ No manual exclusion of objects in this group was necessary. With respect to the three images after 0, 240 and 560 min considered for manual investigation (Figure 3.19), on average 12 % out of all objects can be discarded, referring to a ‘zero bright spot objects’. Furthermore, all other groups comprising objects from one to more than 5 bright spots, contained less than 5 % of objects that were manually excluded. Objects excluded manually might actually belong to another group, show deficient segmentation or are adjacent objects segmented as one single object. With this, the bright spot feature can be considered as a reliable tool for the automatic distinction between single cells, budding cells and cell-clusters, besides enabling the automatic exclusion of artefacts and non-cell objects.

3.4 Conclusions and future perspectives

Advances in microscopy imaging and image analysis as exemplarily demonstrated by means of the oCelloScope instrument offered the possibility for obtaining exciting new insights into a yeast fermentation process. For the very first time, the technology in use was connected continuously to the fermenter via a prototype flow-through-cell. Hence, automatic imaging and image analysis could be performed every 10 minutes over a 10 h yeast (fed-) batch fermentation. In parallel, a second

oCelloScope unit was operated off-line enabling the combination of every 4th on-line image with an image of a diluted off-line sample. This was done in order to overcome the challenge of high biomass concentrations hindering appropriate segmentation of individual cell-objects due to too crowded images. On-line images could be used in particular for the detection of yeast growth, while off-line image data of diluted fermentation samples could be used to yield insights into the morphology dynamics of the growing yeast culture. Advanced algorithms facilitated automatic adjustment of focus and illumination level and thus a microscopic image of the fermentation broth could be acquired within seconds. Moreover, dedicated algorithms allowed for automatic detection of microbial growth. Within this work, OD values between 0.6 and 4 could be followed reliably by means of the BCA algorithm, providing increased sensitivity compared to OD measurements. However, cell concentration referring to an OD value exceeding 4 (a cell dry weight of approximately 2 g/L) were found outside the capacity of image based growth detection by the algorithms applied in this study. OD values below 0.6 were not present during the course of the fermentation but based on publications linked to the technology in use, growth detection below the capacity of OD measurements (OD values < 0.1) with standard spectrometers is possible even with increased sensitivity by this method. Besides, diluted off-line images on samples exhibiting an OD value of 0.1 resulted in exquisite image quality, supporting the assumption of reliable growth detection below the OD values present during the fermentation under investigation in this study. Hence, image based growth detection appears to be especially interesting for the early stages of the microbial growth, and the eventual identifications of contaminations. Note that, for practical reasons the flow-through-cell was chosen in a height of 800 μm , but was available in a height of 300 μm as well. As the BCA algorithm is based on an absorbance (pixel intensity) measurement, a reduction of the sample height will lead to an increase in the biomass concentration detection limit. In the future, the flow-through-cell will be available in a height of 140 μm and thus growth detection up to an OD value of approximately 22 might be possible.

Besides the detection of growth, the morphology dynamics of the growing yeast culture were investigated. It was found that yeast cells undergo a change in morphology with respect to the fraction of single cells, budding cells and cell-clusters present over the fermentation time. In order to facilitate the automatic differentiation between the different cell-objects, the so-called bright spot feature was developed and implemented into the UniExplorer and ParticleTech software. It exploits the fact that yeast cells appear as a bright spot on the images. By counting the number of bright spots associated to a cell-object, the algorithm allows the automatic distinction between single cells, budding cells or a cell-cluster. The bright spot feature was found to work reliably with on average less than 5 % failure and can be used to exclude image artefacts produced by shadings or out-of-focus cells in the background. The latter objects refer to an object with zero bright spots. Within this study, the early stages of (exponential) growth showed a slight shift towards cell clusters, the late stages of (exponential) growth a slight shift towards single cells. However, budding cells were the dominating cell species within this study, investigating the culture over an extended batch phase. Hence, no insight could be gained with respect to a truly challenging cell environment. The

morphological insight was yielded from off-line data of appropriate diluted fermentation samples. The quality of the on-line images was deficient for segmentation due to too high cell concentrations and were in particular challenged by an out-of-focus issue progressing over time. The bottom of the prototype flow-through-cell deformed over the course of the fermentation, so that the images were set more and more out-of-focus. Both, closely adjacent and out-of-focus cell-objects were found to be a crucial challenge for the segmentation algorithms. Closely adjacent cell-objects may result in one big object if the border pixels do touch. Out-of-focus cell-objects led to erroneous segmentation, in other words, cell-objects were not segmented as a human eye would have done it. Hence, segmentation results must be considered very carefully and segmented images must be studied with attention in order to come to the right conclusions. It always must be taken into account that the segmentation results represent the segmentation, and not necessarily the actual nature of the image. Hence, providing appropriate image quality for segmentation is a crucial requirement. By now, ParticleTech Aps has improved the fabrication of the flow-through-cell by means of a mold, overcoming the observed issue of bottom deformation of the flow-through-cell. Moreover, they have developed a dual pump flow controller for automatic dilution. In this way, samples from the fermenter can be automatically diluted providing an appropriate image quality for segmentation. This simplifies tremendously the acquisition of appropriate data. The new flow-through-cell in a height of 300 and 140 μm might well be used to study the morphology dynamics of yeast during different process conditions. First of all, it would be interesting to confirm the first findings by running another yeast batch fermentation considering OD values below 0.1 as a starting point and continuing the image acquisition until cells start to lyse. It might also be interesting to study the effect of relevant process events as, for instance, the failure of aeration / oxygen limitation on the cell morphology. Finally, a yeast production process, as for example relevant for insulin production might be exciting to follow by imaging and image analysis. Eventually, a certain cell morphology might be linked to the production (rate) of the insulin precursor peptide.

When starting this work, the oCelloScope instrument was used off-line and fermentation samples were studied via a 96 well-plate. The need for a flow-through cell was addressed by ParticleTech Aps facilitating a first continuous imaging set-up by means of a prototype flow-through-cell. Challenges linked to the flow-through-cell and the resulting image quality were addressed by a mold for better flow-through-cell fabrication as well as a dual pump flow controller for automatic sample dilution. Thus, by now, a promising and exciting set-up is ready to monitor a yeast-fermentation based on image analysis. Findings might be used to transform the knowledge gained into a novel, image based monitoring strategy, enabling growth detection and moreover the assessment of crucial fermentation parameters based on cell morphology.

3.5 Supplementary Material

Figures S3.1-S3.6 exemplarily represent objects automatically assigned by means of the bright spot feature to the number of bright spots stated below each image. The objects refer to the object segmented on the images acquired off-line on the diluted fermentation samples (1st replicate) at time point 0.

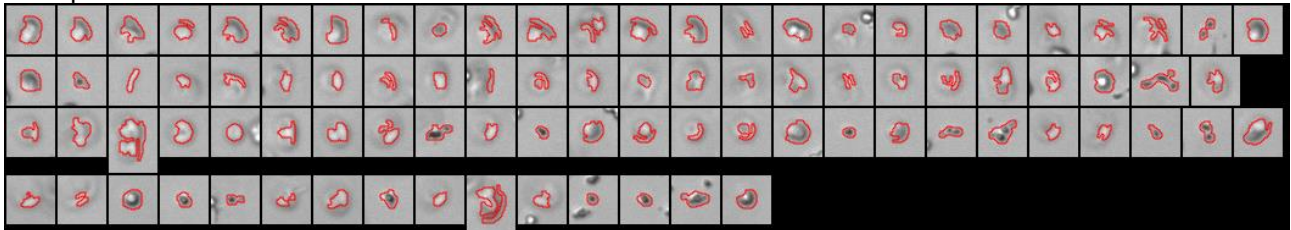


Figure S3.1: Objects with zero bright spots, automatically assigned by the bright spot feature.

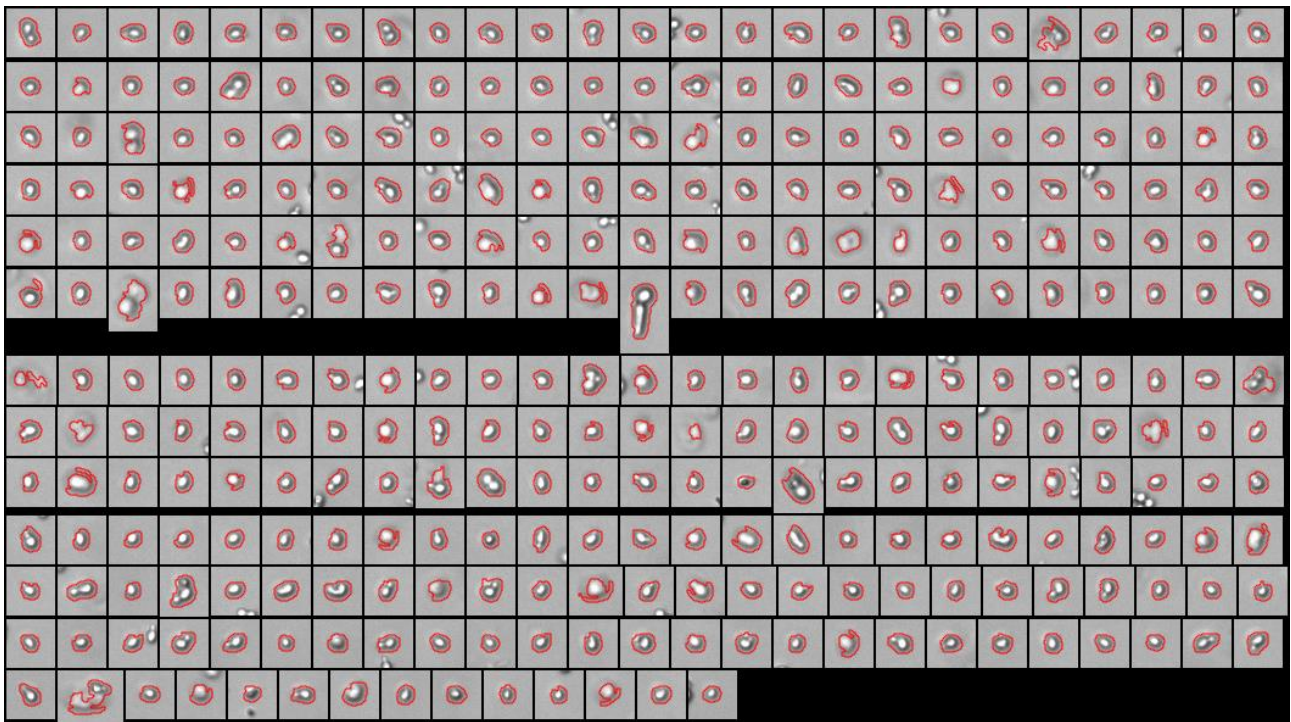


Figure S3.2: Objects with one bright spot, automatically assigned by the bright spot feature.

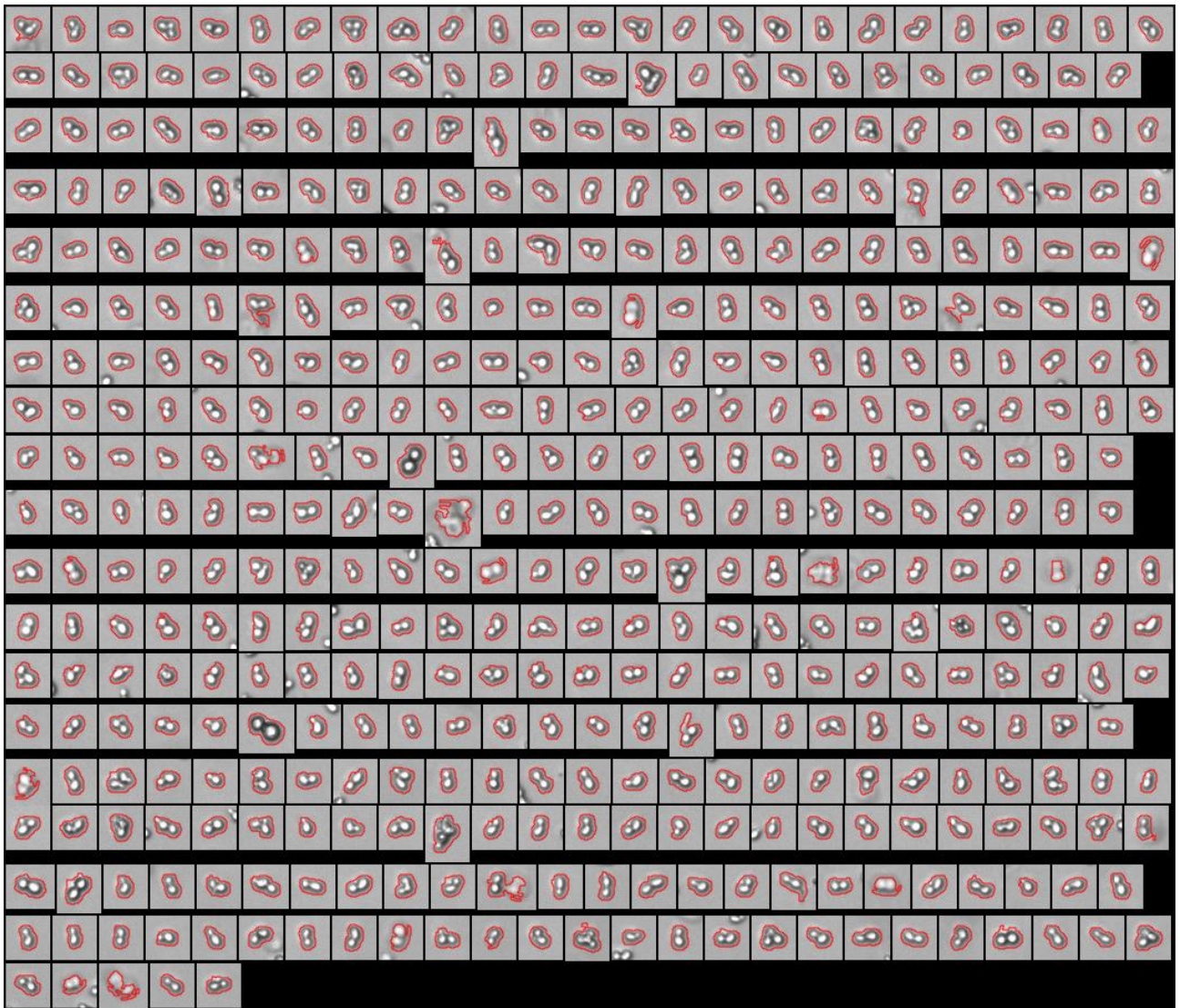


Figure S3.3: Objects with two bright spots, automatically assigned by the bright spot feature.

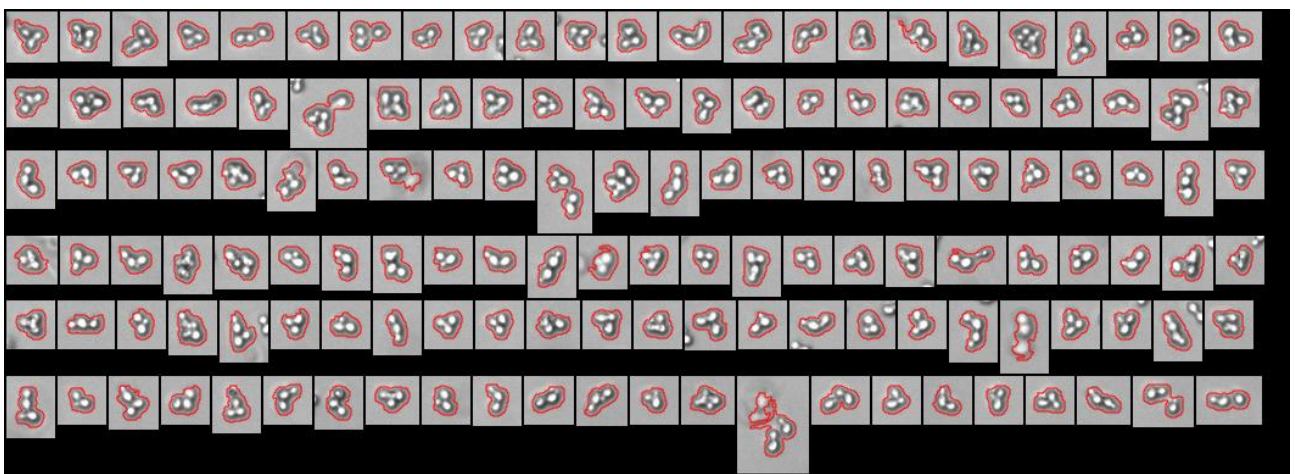


Figure S3.4: Objects with three bright spots, automatically assigned by the bright spot feature.

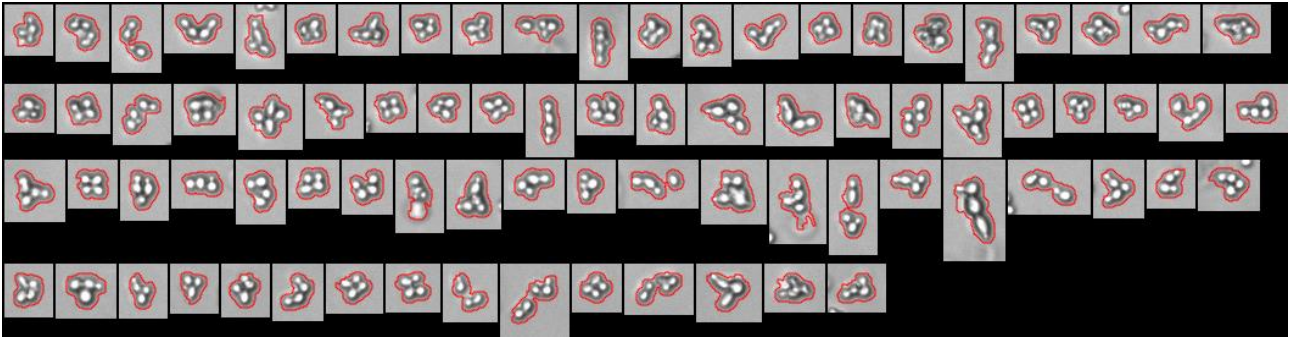


Figure S3.5: Objects with 4 bright spots, automatically assigned by the bright spot feature.

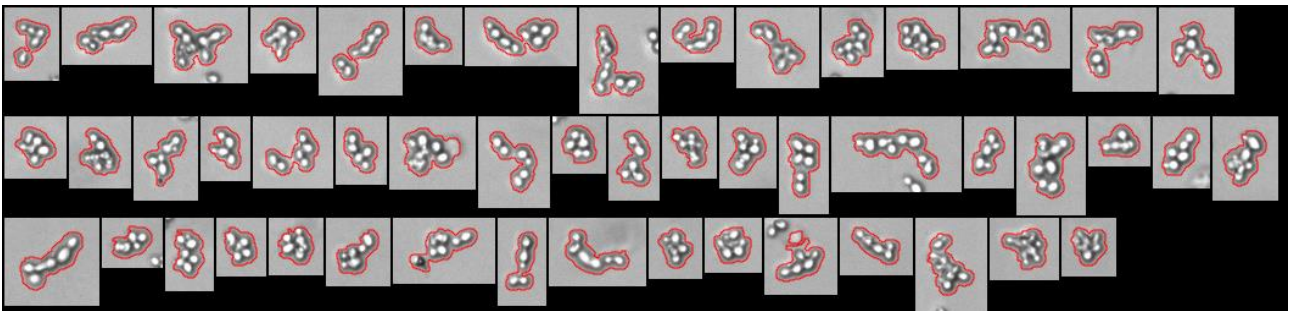


Figure S3.6: Objects with more than 5 bright spots, automatically assigned by the bright spot feature.

Figures S3.7-S3.10 represent objects manually excluded from the groups 1-4 bright spots, as shown in Figures S3.2-S3.5.



Figure S3.7: Manually excluded objects from the group one bright spot (Figure S3.2).

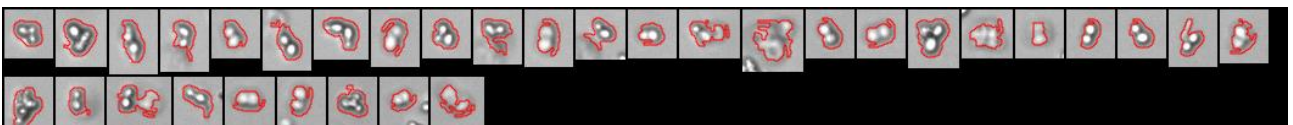


Figure S3.8: Manually excluded objects from the group two bright spots (Figure S3.3).

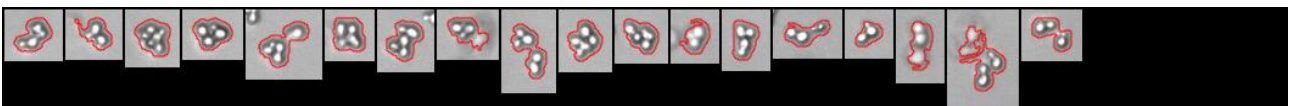


Figure S3.9: Manually excluded objects from the group three bright spots (Figure S3.4).

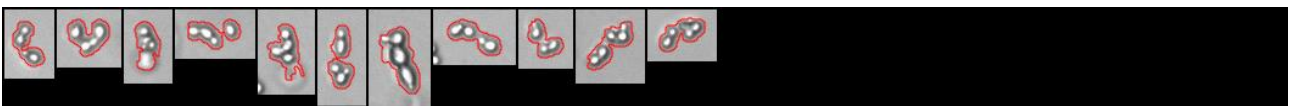


Figure S3.10: manually excluded objects form the group four bright spots (Figure S3.5).

3.6 References

- [1] "Fundamentals of light microscopy and electronic imaging," *Choice Rev. Online*, vol. 51, no. 02, pp. 51-0870-51-0870, Oct. 2013.
- [2] P. P. Mondal and A. Diaspro, *Fundamentals of Fluorescence Microscopy*. Dordrecht: Springer Netherlands, 2014.
- [3] J. Selinummi *et al.*, "Bright Field Microscopy as an Alternative to Whole Cell Fluorescence in Automated Analysis of Macrophage Images," *PLoS One*, vol. 4, no. 10, p. e7497, Oct. 2009.
- [4] A. Ettinger and T. Wittmann, "Fluorescence live cell imaging," in *Methods in Cell Biology*, vol. 123, 2014, pp. 77–94.
- [5] C. L. Rieder, "Mitosis Through the Microscope: Advances in Seeing Inside Live Dividing Cells," *Science (80-.)*, vol. 300, no. 5616, pp. 91–96, Apr. 2003.
- [6] F. Zernike, "Phase contrast, a new method for the microscopic observation of transparent objects part II," *Physica*, vol. 9, no. 10, pp. 974–986, Dec. 1942.
- [7] C. L. Smith, "Basic Confocal Microscopy," in *Current Protocols in Molecular Biology*, vol. Chapter 14, Hoboken, NJ, USA: John Wiley & Sons, Inc., 2008.
- [8] M. M. Frigault, J. Lacoste, J. L. Swift, and C. M. Brown, "Live-cell microscopy - tips and tools," *J. Cell Sci.*, vol. 122, no. 6, pp. 753–767, Mar. 2009.
- [9] R. J. Walter, "Computer-enhanced video microscopy: digitally processed microscope images can be produced in real time," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 78, no. 11, 1981.
- [10] D. L. Coutu and T. Schroeder, "Probing cellular processes by long-term live imaging - historic problems and current solutions," *J. Cell Sci.*, vol. 126, no. 17, pp. 3805–3815, Sep. 2013.
- [11] M. Fredborg, F. S. Rosenvinge, E. Spillum, S. Kroghsbo, M. Wang, and T. E. Sondergaard, "Rapid antimicrobial susceptibility testing of clinical isolates by digital time-lapse microscopy," *Eur. J. Clin. Microbiol. Infect. Dis.*, vol. 34, no. 12, pp. 2385–2394, Dec. 2015.
- [12] L. E. Uggerhøj *et al.*, "Rational Design of Alpha-Helical Antimicrobial Peptides: Do's and Don'ts," *ChemBioChem*, vol. 16, no. 2, pp. 242–253, Jan. 2015.
- [13] T. Kjeldsen, M. Sommer, and J. E. Olsen, "Extended spectrum β -lactamase-producing *Escherichia coli* forms filaments as an initial response to cefotaxime treatment," *BMC Microbiol.*, vol. 15, no. 1, p. 63, 2015.
- [14] N. Ashley, M. Jones, D. Ouaret, J. Wilding, and W. F. Bodmer, "Rapidly derived colorectal cancer cultures recapitulate parental cancer characteristics and enable personalized therapeutic assays," *J. Pathol.*, vol. 234, no. 1, pp. 34–45, Sep. 2014.
- [15] M. Fredborg *et al.*, "Real-time optical antimicrobial susceptibility testing," *J. Clin. Microbiol.*, vol. 51, no. 7, pp. 2047–2053, 2013.
- [16] M. Fredborg, F. S. Rosenvinge, E. Spillum, S. Kroghsbo, M. Wang, and T. E. Sondergaard, "Automated image analysis for quantification of filamentous bacteria," *BMC Microbiol.*, vol. 15, no. 1, p. 255, Dec. 2015.

- [17] P. Tibayrenc, L. Preziosi-Belloy, J. M. Roger, and C. Ghommidh, "Assessing yeast viability from cell size measurements?," *J. Biotechnol.*, 2010.
- [18] V. Camisard, J. P. Brienne, H. Baussart, J. Hammann, and H. Suhr, "Inline characterization of cell concentration and cell volume in agitated bioreactors using in situ microscopy: Application to volume variation induced by osmotic stress," *Biotechnol. Bioeng.*, 2002.
- [19] C. B. Tyson, P. G. Lord, and A. E. Wheals, "Dependency of Size of *Saccharomyces cerevisiae* Cells on Growth Rate," *J. Bacteriol.*, vol. 138, no. 1, pp. 92–98, 1979.
- [20] A.-M. Marbà-Ardébol, J. Emmerich, P. Neubauer, and S. Junne, "Single-cell-based monitoring of fatty acid accumulation in *Cryptocodinium cohnii* with three-dimensional holographic and in situ microscopy," *Process Biochem.*, vol. 52, pp. 223–232, 2017.

Chapter 4

Monitoring of microbial growth by the Cell-Growth-Quantifier

Preface

As outlined in chapter 3, the monitoring of biomass is crucial to any fermentation process. No other parameter demonstrates the progress of the fermentation process in such a way as the increase in biomass itself. However, the detection of biomass is still compromised by off-line analysis and probes insensitive to the entire fermentation process. The following chapter introduces the possibility of continuous, non-invasive monitoring of microbial growth by means of a light-backscatter sensor, called Cell-Growth-Quantifier (CGQBIOR, aquila biolabs GmbH, Germany). The sensor is able to detect the microbial growth via the fermenter wall (in case of a glass vessel) or a small glass window in the vessel (in case of a stainless steel vessel), respectively. The development of this technology was originally dedicated for accurate and high speed data acquisition in shake flasks under shaking conditions. For this purpose, the technology has become more and more popular, as exemplarily demonstrated by the integration of the CGQ technology into the Infors product portfolio in October 2017. The sensor design has been modified towards the application to bioreactor vessels (CGQBIOR). However, the application towards fermentation vessels is rather new. The sensor is especially attractive due to its simple to use and non-invasive application, facilitating the on-line detection of microbial growth over the entire fermentation course. The following chapter is dedicated to the investigation of the sensor performance under relevant process conditions. It presents the backscatter measurement technology and its application towards a yeast fermentation process in a single-wall, glass lab-scale fermentation vessel (Applikon).

I decided to implement this case study into the thesis when I met aquila biolabs during my stay at the 5th BioProScale Symposium, hold in Berlin, Germany, in March 2018. I would like to acknowledge Julius Netzer (aquila biolabs, Germany) for the inspiration towards this technology and open communication and helpful feedback when conducting the experimental work.

4.1 Introduction

Generally, biomass probes rely on spectroscopic (optical density, IR, Raman, Fluorescence spectroscopy), optical (in-situ microscopy, image analysis) and electrochemical (impedance spectroscopy) techniques [1], [2], [3], [4], [5]. The CGQ belongs to the first class. Essentially, the spectroscopic techniques differ in terms of the wavelength used and the geometric arrangement of the optical pathway, defined *i.a.* as the angle between light source and detector. The traditionally

used turbidity measurement (optical density, OD) performed off-line determines the absorbance of a sample with a constant light path at a 0° angle. The light source and detector are facing each other and a sample is placed in between. Absorbance measurements are restricted to a rather small linear range, and they do not facilitate biomass detection covering the complete bioprocess, from inoculation ($OD < 0.2$) to harvesting ($OD > 30$) without sample preparation. OD measurements are only sensitive towards the entire bioprocess when properly diluted. Measuring scattered light instead of absorbance can be considered as another form of turbidity measurement, however, with a largely extended linear range. Besides, light scattering is a function of wavelength and angle, allowing for adjustment with respect to lower and higher cell concentration levels. Higher wavelengths produce higher scattering than lower wavelength. The CGQ facilitates the monitoring of biomass in standard and high cell density processes with two different LED as light source at a fixed scattering angle of 180° . An arrangement of 90° is more sensitive towards lower cell concentrations, while a 180° geometry between light source and detector facilitates the detection of rather high cell concentrations [2], [6]. Hence, the CGQ geometry allows for the detection of high cell concentrations and for a design, in which light source and detector are placed next to each other, not in front of each other. Besides, backward scattered light (angles $> 90^\circ$), is more sensitive towards the cell membrane structure [7], potentially allowing the tracking of morphological changes. The CGQ was developed with the particular scope of on-line biomass detection over the whole process range and not for tracking morphological changes. However, differences in the slope of calibration curves correlating OD measurements to the backscatter intensity are species dependent and this is explained by different cell morphologies [8]. Within this work, the application of the CGQ technology is focused on the detection of yeast growth. It was tested with respect to the influence of operational changes, known to affect several (in-situ) sensor signals. This included the adjustment (or failure) in stirrer speed and aeration rate, solids in the fermentation broth, differences in the fermentation medium (transparent vs. colored) and process events as substrate addition (feeding). It is of no doubt that a reliable, online biomass signal obtained non-invasively is of clear benefit for any fermentation process. Note that, on-line biomass monitoring is a great deal and not yet established in research and industry.

4.2 Materials and Methods

4.2.1 Light scattering

Light scattering in a solution (suspension) of particles can be understood as a change in photon direction, when the photon collides with a small object [9]. In a fermentation process, such a small object may be a particle, a cell, a micelle, an air bubble or larger biopolymers (> 1 kDa) as peptides, enzymes and proteins. Theoretically, also the fermenter's inner equipment as stirrer, baffles, sampling port and aeration pipe might interfere (reflect) with the light and contribute to the amount of light that is scattered or respectively, reflected back towards the detector. Hence, the position of the sensor across all experiments was fixed to avoid baseline drifts.

The interaction of light and particle will cause the light beam to be attenuated when passing through the solution. The light can be attenuated by scattering or absorption processes. In both cases, the light intensity after the solution will decrease exponentially with the path length x (see Figure 4.1) through the solution. Light attenuation based on absorption and scattering is schematically illustrated in Figure 4.1.

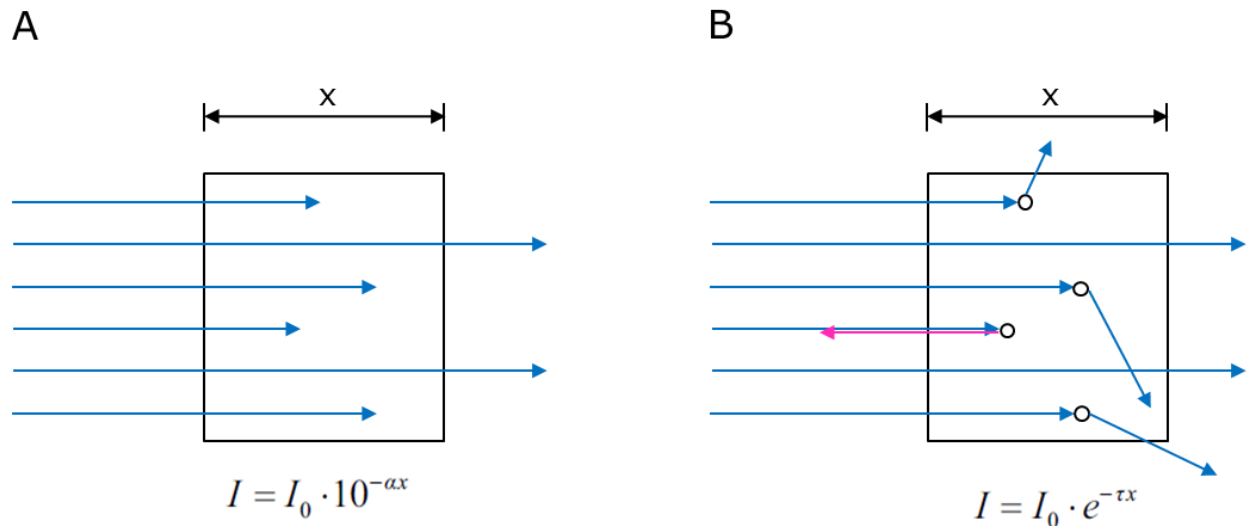


Figure 4.22: Decrease in light intensity based on either absorption (A) or scattering (B). The intensity of the light I after the solution is described as a function of the incoming light intensity I_0 and the absorption coefficient (α) or, respectively, the turbidity coefficient (τ). The two different bases (10 and e) for the exponential decays are a matter of convention. In Figure 4.1B, 180° scattered light is indicated as a pink arrow.

The larger the particles and the higher the concentration, the higher the scattering. It is interesting to notice that the particle size is reflected in the scattering angle. If the detector measuring the scattered light is mounted on a so –called goniometer, controlling from which angle the scattered light is recorded, it is possible to obtain information about the size of the particles (molecules), and if the concentration is known, about the molecular weight. [10].

4.2.2 The CGQ

The CGQ measures turbidity based on the intensity of the light scattered back by the cells exhibiting a scattering angle of 180°. The scattering angle ensures that the light scattering is caused by the cell membranes [11], [12]. Consequently, the increase in the number of cells results in an increase in back-scattered light. The increase in backscatter intensity is exploited here while conventional turbidity measurements (OD) detect the decrease in light intensity due to scattering, reflection and absorption by the cells. Each CGQ sensor plate consists of a combination of light source (LED) and photodiode, mounted next to each other. LED and photodiode are spatially separated to prevent a direct illumination of the detecting part. The CGQ facilitates backscatter measurements for standard (OD levels between 0.5 and 50) and high-cell-density (OD levels between 15-300) fermentation processes, applying a LED of 521 nm and 940 nm, respectively. The sensor is integrated in a foam housing and mounted directly onto the wall of the reactor by means of a flexible belt. The process (glass) window

required for its application must be approximately 1.5 cm x 2 cm. The sensor, its mount to the reactor and measurement principle are illustrated in Figure 4.2.

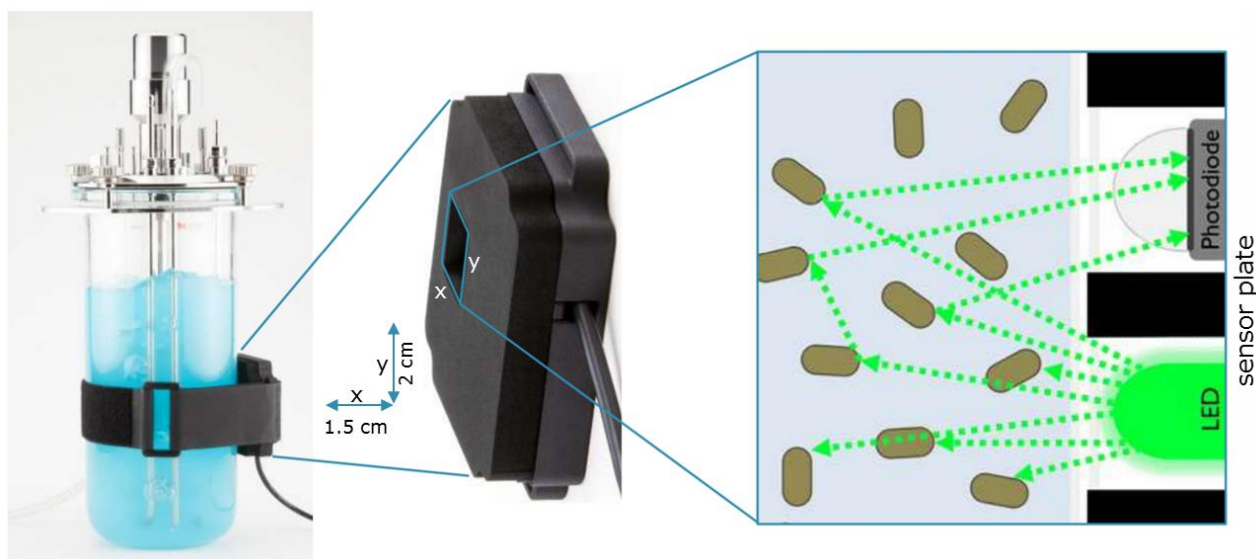


Figure 4.2: From left to right: Mounting of the sensor on the reactor by means of a flexible belt, sensor housing and detection principle. The size of the detection window is indicated by x and y. The green LED exemplarily demonstrates the standard configuration (at 521 nm) [13].

The physical set-up to the fermenter is quick and easy ('plug and play'). As the CGQ works non-invasively, it can be installed / uninstalled any time during the fermentation and no sterilization or extensive cleaning is required. The sensor can be operated at process temperatures between 10 °C and 80 °C, covering conventional process temperatures applied during fermentations. The data streams from up to 16 fermenters are bundled at a base station and forwarded to a local computer via USB where it is presented to the user by means of the CGQuant software. The CGQuant software handles the control of the CGQ and all data analysis, including charting of online data, data documentation, annotation of process events, inclusion of calibration files (for *e.g.* OD or cell dry weight calibration), the calculation of growth rates, and statistical analysis of replicates.

Currently, three other devices serving as a technical benchmark for the CGQ with respect to determination of microbial growth in shake flasks are available (the SFR Vario (Presens GmbH, Germany), the OD-Monitor (TAITEC, Japan) and the OD-Scanner (BugLab LLC, USA). However, they demand a static measurement set-up (meaning that the shaking must be stopped for their application) and don't offer a solution that can be applied to a bioreactor [8]. Note that, the CGQ was developed for shake flasks and its particular scope is defined by continuous monitoring over an entire shake flask cultivation process under shaking conditions. The application to fermentation reactor vessels is new and not extensively studied thus far.

4.2.3 Experimental conditions tested by the CGQ

4.2.3.1 CGQ set-up

The experiments were conducted with the CGQBIR, the CGQ model designed for the application to fermentation reactor vessels. In the following, the CGQBIR is simply called CGQ. The sensor was equipped with a 521 nm LED, facilitating biomass measurements between OD values of 0.5 and 50, as stated by the company. No reference organism was mentioned. The measurement mode was set to external light compensation. For testing the effect of changes in operational conditions, the data collection rate was set to 2 seconds. For the fermentation processes, the data collection rate was set to 2 min. The CGQ was mounted in the middle of the reactor vessel, assuring that no inner parts were directly in front of the sensor window and the position was marked to avoid signal variation based on positioning.

4.2.3.2 Operational conditions under investigation

The environment inside the fermenter is highly effected by operational events. Operational events as feeding (substrate or acid and base for pH control) as well as adjustment of aeration rate and stirrer speed change the environment inside the bioreactor physically. Such operational events are generally known to cause disturbances in several sensor signals. Hence, the influence of stirrer speed and aeration rate on the sensor signal was tested. Additionally, the effect of medium composition (water and YP (yeast extract and peptone) medium, outlining a transparent and a colored medium, respectively, and the presence of particles (YP medium containing potato protein powder) was investigated. Particles as potato protein powder are part of certain complex media. They are enzymatically hydrolyzed prior to the fermentation process to release amino acids into the broth. The potato protein remains inside the fermentation broth as a nutritional source during enzyme production processes, in which the target enzyme is active on the remaining potato protein particles, thus degrading it more and more in the course of the fermentation. The amino acids released are in turn important building blocks for the expression of the enzyme itself. As optical density is not selective towards cells but towards all particles in solution, the presence of particles does cause practical issues to OD measurements, depending on the particle concentration. If the particle concentration in the broth is high, the detection maximum of most spectrometers is already reached by the particles themselves, making the OD measurement insensitive to a change in microbial biomass. Dilutions may be considered to circumvent the high particle concentrations but according results are often error prone and of poor reproducibility. The degradation of particles during the process would impose another factor to account for. However, the detection of microbial growth in the presence of particles is a valuable advantage of the CGQ over OD measurements.

Conditions tested are outlined in [Figure 4.3](#). The stirrer speed was varied between 400 rpm and 1100 rpm, the aeration rate was varied between 0.5 vvm and 2 vvm. For the effect of the stirrer speed, a signal was collected at 1vvm for 10 min, for each stirrer speed (conditions along the red arrow in

Figure 4.3). 1 vvm represents the operational aeration rate and a change in stirrer speed may represent a case where the dissolved oxygen tension is maintained above a certain set point (>30 %) by a stirrer speed cascade. When studying the effect of the aeration rate, each stirrer speed was tested over a time period of 2.5 min. The conditions tested with respect to aeration rate are represented by set-points along the green arrows in Figure 4.3. A change in stirrer speed or aeration rate might not necessarily represent a control action, but eventually a failure in the power supply or the aeration system. Aeration is easily affected by a decrease in air supply pressure or a pressure increase inside the reactor, due to a blocked off-gas filter.

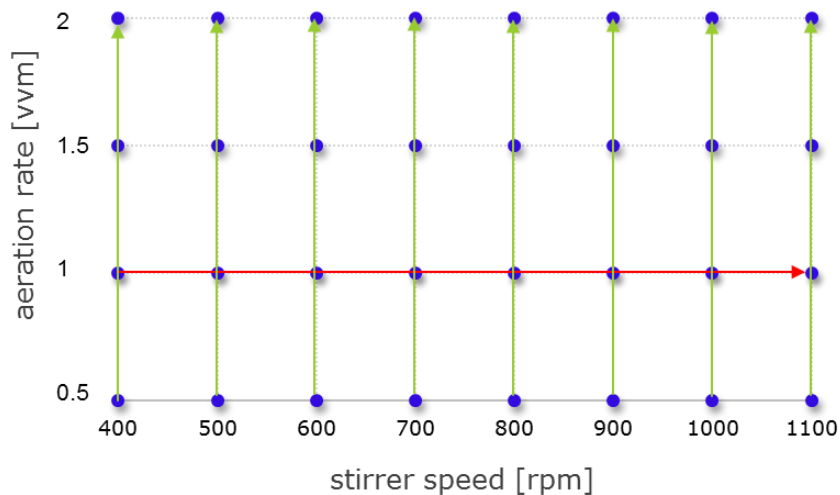


Figure 4.3: Experimental design of testing conditions. The green arrows indicate the conditions tested at fixed stirrer speed. The red arrow indicates the conditions investigated at fixed aeration rate.

The experimental design was performed in all three media under investigation.

4.3.2.2 Fermentation conditions under investigation

Finally, aiming at the investigation of biomass monitoring performed by the CGQ, two standard yeast fermentations in YP medium as described in detail in chapter 2.2.1.1 Yeast fermentation processes, were performed. Fermentation 2 contained a constant particle concentration of 50 g/l. Both fermentations were subject to glucose feeding in order to challenge the sensor with higher biomass concentrations. The feeding solution for both fermentations contained a glucose concentration of 500 g/l. During the particle-free fermentation (fermentation 1), the CGQ data was complemented hourly with off-line OD and HPLC data for glucose, ethanol, glycerol and acetic acid, and furthermore every 2nd hour with off-line dry weight measurements. Fermentation 2, containing particles, was supplemented with online off-gas data (CO₂), as OD and dry weight measurements were not possible to perform reliably. OD, dry weight and HPLC measurements were performed as described in chapter 1.2.3 and 2.2.3.1. Off-gas measurements were provided by means of a prototype of the NxPAS gas analyzer, shortly introduced in chapter 3.2.4.

4.3 Results and Discussions

4.3.1 Influence of stirrer speed and aeration rate on the sensor baseline signal

The baseline of the CGQ in different media (water, YPD medium and YPD medium containing 50 g/l potato protein particles) was tested for stirrer speeds between 400 rpm and 1100 rpm as well as for aeration rates between 0.5 vvm and 2 vvm, respectively (Figure 4.3). The results are presented per medium under investigation.

4.3.1.1 Water

The change of stirrer speed at a constant aeration rate of 1 vvm is presented in Figure 4.4, the influence of aeration rate at different stirrer speeds is demonstrated in Figure 4.5.

Change in stirrer speed at constant aeration rate (Figure 4.4)

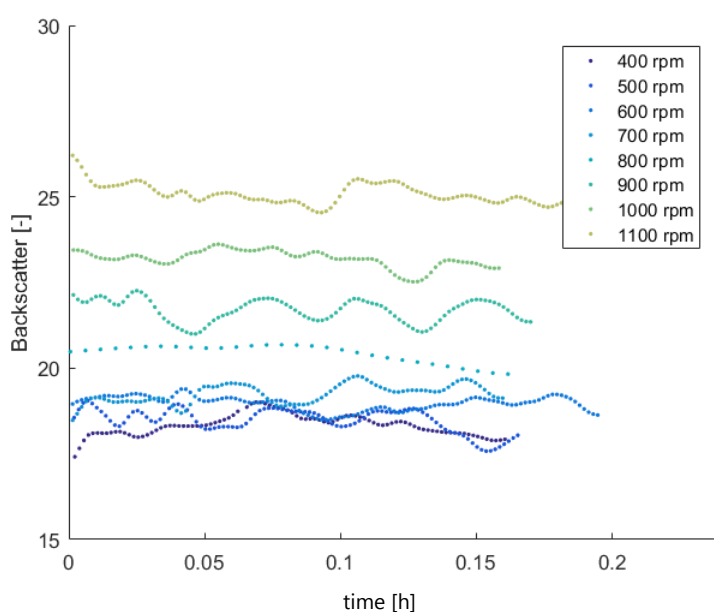


Figure 4.4: Variation in the CGQ background signal at a constant aeration rate of 1 vvm, as a function of the stirrer speed as indicated in the legend.

Stirrer speeds between 400 rpm to 1100 rpm (in steps of 100 rpm in between, see experimental design in Figure 4.3) resulted in measured backscatter values between 17 and 26.

Within the range of 400 rpm – 700 rpm, the level of the background signal could be considered as independent of the stirrer speed. From 800 rpm to 1100 rpm, the background signal was clearly increasing with increasing stirrer speed and the increase seems to follow a linear relationship as a function of stirrer speed (Figure 4.4). The increase in the background signal from 20 to 25 when increasing the stirrer speed from 800 rpm to 1100 rpm represents a signal increase of around 25 %. Assuming that the water was free of any particles and larger molecules, the backscatter intensity detected was the result of metal parts inside the bioreactor and air bubbles. The observed effect can be explained by the dispersion of incoming air by the stirrer. At stirrer speeds below 800 rpm, the dispersion of air can be considered as rather poor and large air bubbles escaped to the reactor

headspace. These air bubbles may just let the light through, and the scattering is minor. At stirrer speeds above 800 rpm, the dispersion of air bubbles is effective and it is expected that the size of air bubbles decreases further with increasing power input (stirrer speed). The dispersed air bubbles can be considered as small particles inside the liquid. The decrease in air bubble size comes along with an increase in air bubble (particle) concentration in the detection area, thus increasing the sensor signal.

Change in aeration rate at constant stirrer speed (Figure 4.5)

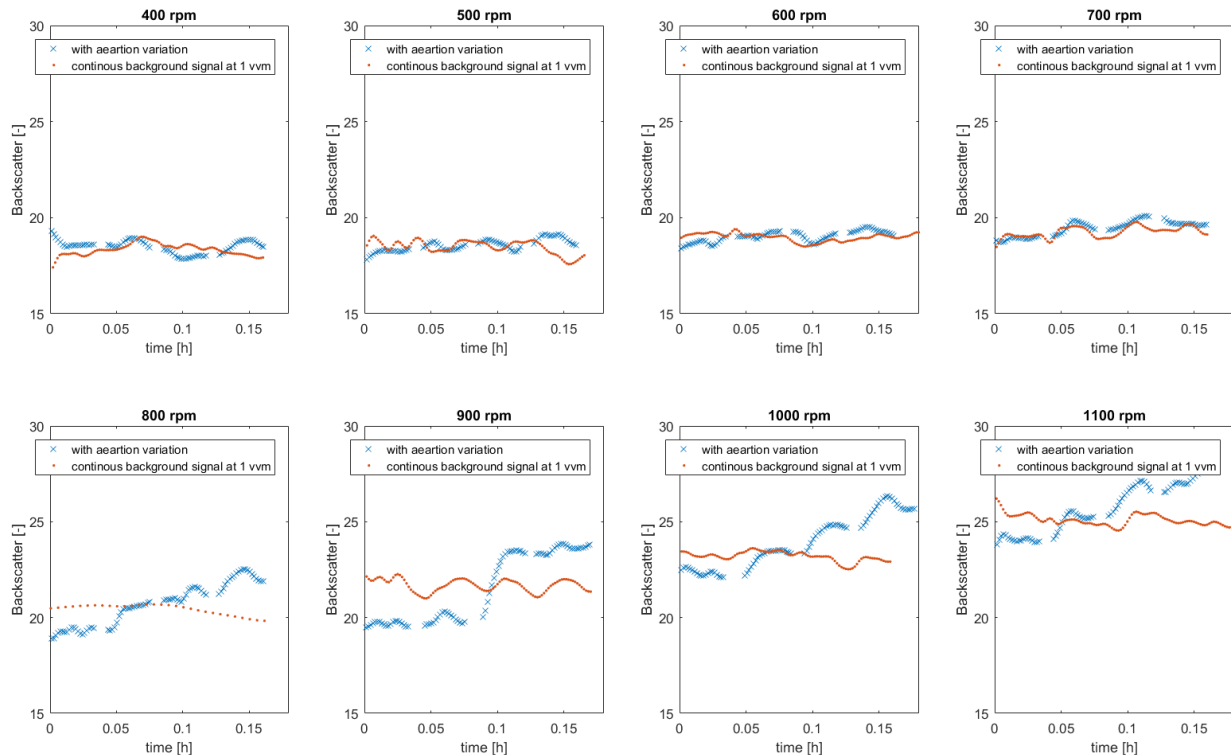


Figure 4.5: Variation in the CGQ signal at constant stirrer speed as a function of the aeration rate (blue crossed line). Starting with 0.5 vvm at every stirrer speed, the aeration rate was increased every 2.5 min in steps of 0.5 (see experimental design in Figure 4.3). The stirrer speed is noted above each graph and the gap in the blue signal marks the change in aeration rate, when no signal was recorded. The red signal represents the backscatter intensity at 1 vvm collected over 10 min for each stirrer speed, providing a visual reference.

A change in aeration rate did not affect the baseline when applying stirrer speeds from 400 rpm to 700 rpm. At stirrer speeds above 800 rpm, the signal increased with an increase in aeration rate (Figure 4.5). The observation can be explained by the higher air flow passing into the reactor. The more air enters the reactor at higher stirrer speeds, the more air can be effectively dispersed resulting in an increase in air bubble (particle) concentration inside the detection area.

4.3.1.2 YP medium

Change in stirrer speed at constant aeration rate (Figure 4.6)

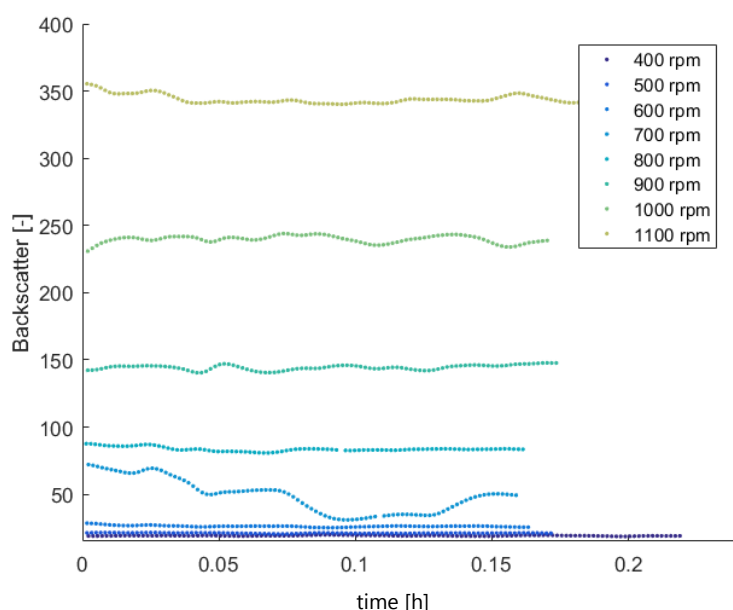


Figure 4.6: Variation in the CGQ backscatter signal at a constant aeration rate of 1 vvm, as a function of the stirrer speed as indicated in the legend.

Generally, the backscatter intensity increased with increasing stirrer speed (at a constant aeration rate of 1 vvm), as observed for water as medium. However, the backscatter intensity (signal) observed in YP medium was generally higher than in water. Stirrer speeds between 400 rpm and 1100 rpm resulted in backscatter values of 25 - 350, while the highest backscatter intensity obtained in water was 27. This might be explained by a lot of larger biopolymers, essentially peptides of different sizes, present in the YP medium. The signal increased by more than 250 % when increasing the stirrer speed from 800 rpm to 1100 rpm and the relation appears to be rather exponential. This represents a 10 fold higher signal increase than in water, at respective stirrer speeds. As in water, the increase in background signal was most pronounced at stirrer speeds between 800 rpm and 1100 rpm. The effect of the stirrer speed on the backscatter intensity for stirrer speeds above 700 rpm was significantly higher than for the respective conditions in water. It is assumed to be explained by the fact, that the YP medium is prone to strong foaming while water is not. The physicochemical properties of YP medium differ essentially from water, as a lot of different biomolecules and ions are dissolved in this complex medium. The physicochemical properties of the medium, such as the surface tension, affect the dispersion of air. The surface tension of water is higher than the surface tension of most other liquids, due to the hydrogen bonding between the water molecules. Hence, air bubbles in water are unstable, due to the high surface tension of this liquid. The net of hydrogen bonding is affected in the YP medium, as a lot of large biomolecules had to be incorporated and dissolved. As a result, the surface tension of YP medium is significantly reduced compared to the surface tension of water. The reduced surface tension is the result of the dissolved biomolecules, acting as surfactants and thus, stabilizing the air bubbles. Hence, the average air bubble size in YP

medium is expected to be larger than in water. Larger particles (air bubbles) cause higher scattering, explaining the increased sensor signal in YP medium compared to water.

Change in aeration rate at constant stirrer speed (Figure 4.7)

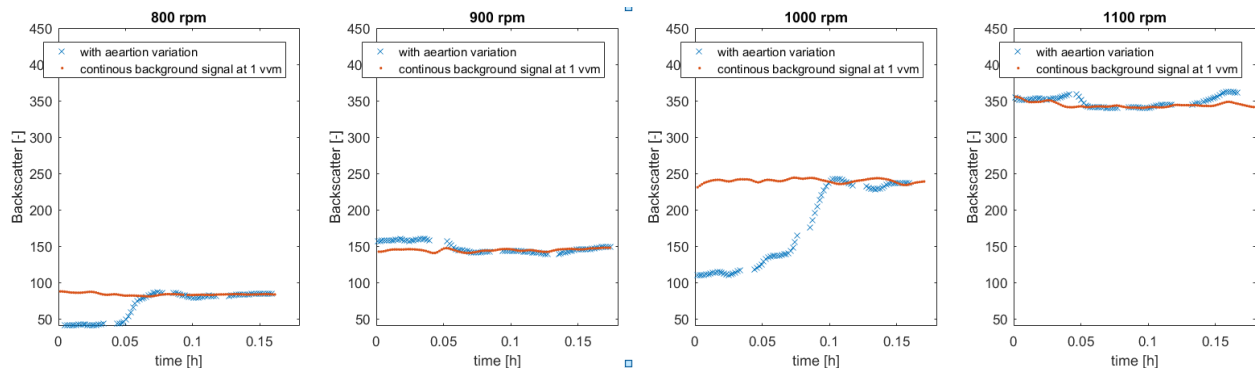


Figure 4.7: Variation in the CGQ signal at constant stirrer speed as a function of the aeration rate (blue crosses). Starting with 0.5 vvm at every stirrer speed, the aeration rate was increased every 2.5 min in steps of 0.5 (see experimental design in Figure 4.3). The stirrer speed is noted above each graph and the gap in the blue signal marks the change in aeration rate, when no signal was recorded. The red signal represents the backscatter intensity at 1vvm collected over 10 min for each stirrer speed, providing a visual reference.

It was observed in water and in YP medium, that the influence at stirrer speeds below 800 is minor, due to inefficient dispersion of air bubbles. Besides, 800 rpm and above, representing the event of dissolved oxygen control by a stirrer cascade, were the relevant process conditions with respect to biomass monitoring in the two fermentation processes. Hence, the effect of the aeration rate in YP medium containing particles was only tested at stirrer speeds ranging from 800 rpm to 1100 rpm. For 900 and 1100 rpm, the aeration rate did not affect the CGQ signal. For 800 rpm and 1000 rpm, a steep increase in the sensor signal was observed when changing to an aeration rate of 1 vvm and 1.5 vvm, respectively. After the steep increase, the signal stayed stable for 1.5 vvm and 2 vvm and a further increase in aeration rate did not affect the signal again. This might be explained by the particular fluid dynamics during the different operations and the results should be reproduced in order to confirm these findings.

4.3.1.3 YP medium containing particles

The presence of particles was simulated by potato protein powder, in a concentration of 50 g/l.

Change in stirrer speed at constant aeration rate (Figure 4.8)

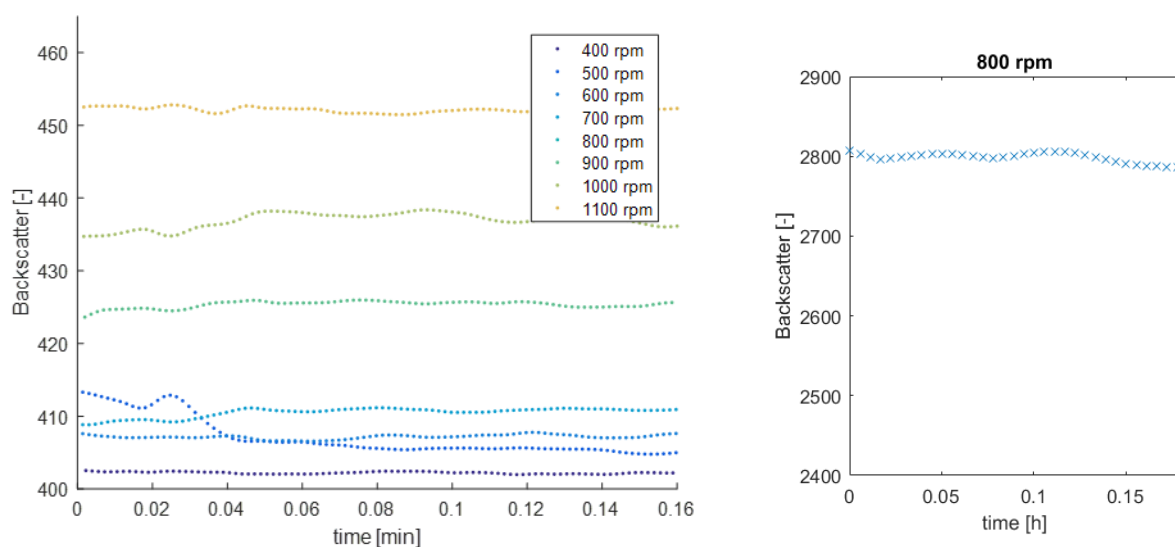


Figure 4.8: Variation in the CGQ background signal at a constant aeration rate of 1 vvm, as a function of the stirrer speed as indicated in the legend. The signal at 800 rpm must be considered as an experimental outlier as it reached backscatter intensities of around 2800 (shown on the right).

For the stirrer speeds under investigation, the backscatter intensity yielded baseline values of around 400 at 400 rpm and around 450 at 1100 rpm. As observed before, the backscatter signal increased with increasing stirrer speed and the effect was most pronounced from 700 rpm to 1100 rpm. Within this range, the increase in base line accounted for approximately 25 %. The phenomenon can be explained as before, by a higher air bubble concentration inside the detection area yielded by a better dispersion of the incoming gas by a higher impeller speed. It is very interesting to notice though, that the extent of this effect is highly dependent of the medium itself. Particles seem to 'stabilize the system', as a change in stirrer speed in the presence of particles resulted in a significantly lower signal increase (12.5 %), as observed for YP medium (250 %). When solid particles were present, changes in air bubble size and concentration clearly affected the CGQ signal to a lesser extent. It might be considered that the present particle concentration was close to the upper detection limit and thus, the sensitivity towards changes in particle concentration caused by air bubbles was less than without particles. A backscatter intensity of 3000 was measured for an experimental outlier, suggesting that the sensor's upper detection limit is far above the measured 450 (intensity units). However, it is of course not a proof.

Change in aeration rate at constant stirrer speed (Figure 4.9)

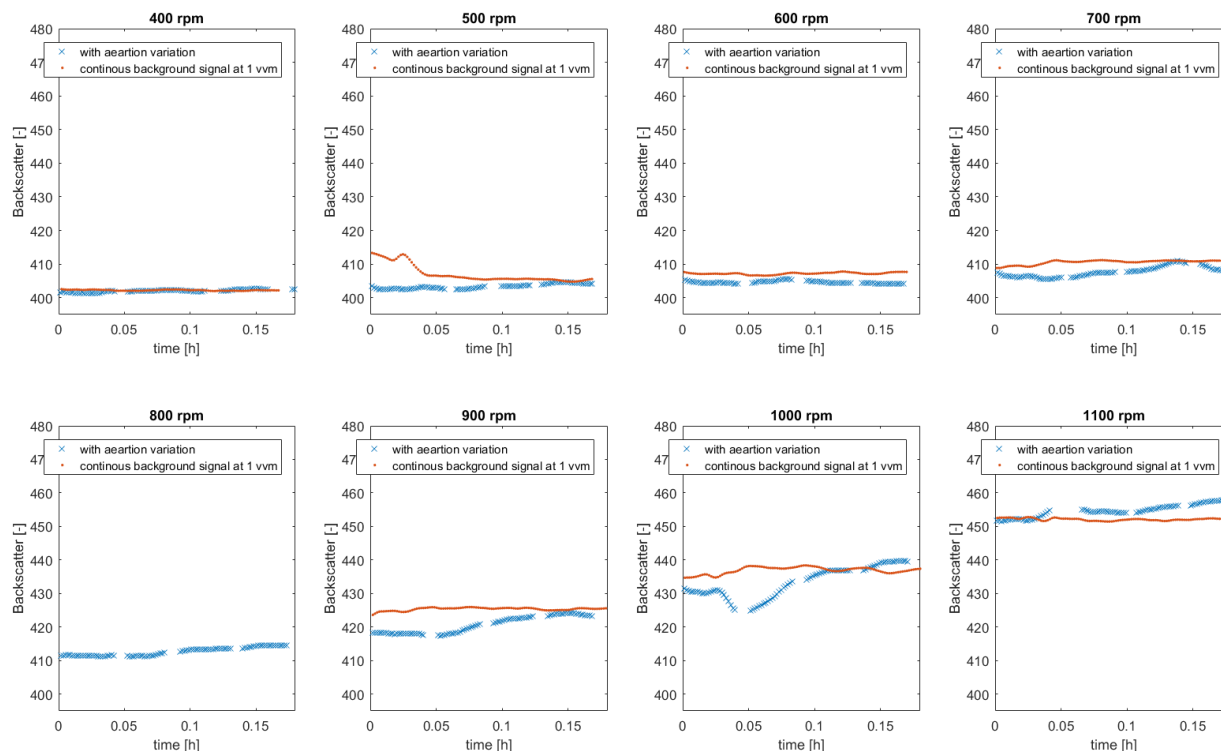


Figure 4.9: Variation in the CGQ signal at constant stirrer speed as a function of the aeration rate (blue crossed line). Starting with 0.5 vvm at every stirrer speed, the aeration rate was increased every 2.5 min in steps of 0.5 vvm (see experimental design Figure 4.3). The stirrer speed is noted above each graph and the gap in the blue signal marks the change in aeration rate, when no signal was recorded. The red signal represents the backscatter intensity at 1 vvm collected over 10 min for each stirrer speed, providing a visual reference.

Generally, the change in aeration rate at stirrer speeds between 400 and 1100 rpm can be considered as insignificant. When increasing the aeration rate, a small increase in the CGQ signal (around 2 %) could be observed for stirrer speeds above 700 rpm. However, this can be considered as negligible compared to an observed increase of 50 % (at 800 rpm) to 150 % (at 1000 rpm) under the same conditions in YP medium without particles.

4.3.1.4 Influence of stirrer speed and aeration rate – summary

It can be concluded that the stirrer speed and the aeration rate clearly affected the backscatter intensity. Generally, the higher the stirrer speed and the higher the aeration rate, the higher the CGQ signal. The signal increase was most pronounced at higher stirrer speeds (> 700 rpm) and might be neglected at lower stirrer speeds (< 700 rpm). The increase in sensor signal with increasing stirrer speed or aeration rate, respectively, can be explained by an increased air bubble concentration in the detection area. At stirrer speeds above 700 rpm, the dispersion of incoming air is effective. With increasing power input, the bubble size decreases while the bubble concentration increases, hence resulting in an increased sensor signal. Increasing the stirring rate affected the sensor signal to a higher extent than an increase of the aeration rate. In particular the effect of stirring was found to be

highly dependent of the medium. It was found strongest in YP medium (an increase of 150 % was observed) and lowest in YP medium containing a particles (an increase of 10 % was observed) when increasing the stirrer speed from 700 rpm to 1100 rpm. In water, the signal increased by around 25 % when increasing the stirrer speed from 700 rpm to 1100 rpm. The differences can be linked to the different physicochemical properties of the medium affecting the dispersion of the incoming air. The reduced surface tension of YP medium compared to pure water is assumed to explain the strong effect in YP medium. In YP the YP medium, the signal increase due to an increase in stirrer speed was significantly reduced in the presence of potato protein particles.

These results clearly reveal that a change in sensor signal might be caused by operational events such as a change in stirrer speed and aeration rate. These events need to be properly tracked in order to interpret the signal correctly. The observed trends seemed to undergo a linear or exponential relationship and hence the backscatter intensity might be corrected by the respective factor.

4.3.2 Monitoring of Biomass by the CGQ

4.3.2.1 Fermentation 1

A standard yeast (fed-) batch fermentation was performed over 30 hours and the growth of biomass was followed online by means of the CGQ. The data was complemented every hour by off-line OD and HPLC measurements and every 2nd hour by off-line dry weight (DW) measurements. Glucose was fed manually three times in a volume of 100 ml (concentration of 500 g/l) in order to prolong the exponential growth phase. The stirrer speed was constant at 800 rpm throughout the process. The fermentation progress followed by the CGQ, OD and dry weight is shown in [Figure 4.10A](#). Calibration curves for the correlation of the OD and dry weight to the backscatter signal are shown in [Figure 4.10 B](#) and [4.10 C](#). The CGQ detected the increase in biomass in good agreement with OD and dry weight data. During the fermentation, the backscatter intensity increased from 50 to 450. The initial value of 50 is below the value previously obtained in YP medium at 800 rpm (around 90, see [Figure 4.6](#)). This can be explained by supplementing the raw (YP) medium with 100 ml of a glucose solution (400 g/l) at the beginning of the fermentation process. The addition of the glucose solution diluted the YP medium, hence decreasing the sensor signal. The signal at 800 rpm before and after glucose addition can be found in the supplementary material, [Figure S 4.1](#). The fermentation yielded a final dry weight value of 20 g/l, corresponding to an OD value of 55.

The whole fermentation process was continuously followed by the CGQ. No detection limit was observed during this fermentation. Process events such as the feeding of glucose (indicated in [Figure 4. 10 A](#) as black arrows) caused a small signal variation, generally characterized by a small drop in signal, due to a local dilution of the detection area during feeding.

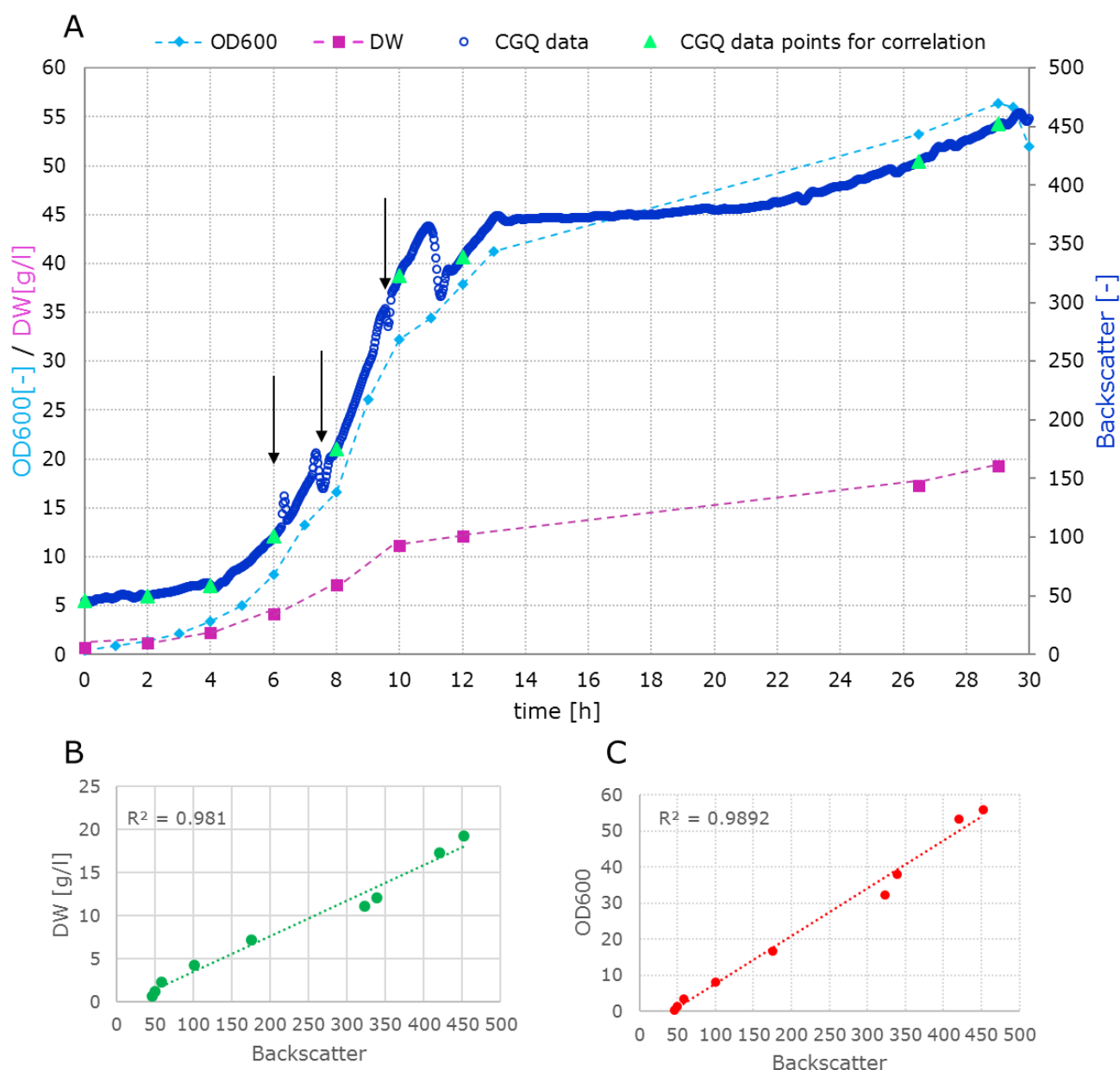


Figure 4.10: A) Biomass increase during a 30 h yeast fed-batch fermentation followed continuously on-line by the CGQ as backscatter intensity. The data was supplemented with off-line OD and dry weight (DW) measurements. Glucose addition (100 ml of a 500 g/l glucose solution) is indicated with black arrows. The signal drop after 11 h was not related to an active process operation. It might relate to a drop in air pressure sometimes happening during the fermentation due to a pressure drop in the supply line. The data points in the backscatter signal used for the correlation with the OD and the DW are marked with a green triangle. B) Correlation between the DW and the backscatter intensity. C) Correlation between the OD and the backscatter intensity.

Feeding took place manually by means of a sterile syringe through the feeding septum, located at the top of the reactor, with the 'feeding stream' entering close to the position of the sensor. The signal drop after 11 h could not be related to an active operational event, but might be explained by a drop in aeration rate which sometimes happens when the supply pressure decreases briefly. Good OD and DW correlations were obtained with a correlation coefficient of greater than 0.98 (see Figure 4.10 B and 4.10 C). Hence, the calibration curves can reliably be used to convert the on-line

backscatter data into OD or dry weight data, if preferred. Biomass dilution due to acid, base and glucose addition was not considered here. No samples for reference analysis were taken over night. However, the diauxic shift from glucose to ethanol consumption after 13 h was clearly visible in the sensor signal. It is indicated by a decrease in exponential growth rate. The CGQ data was in agreement with the HPLC data, confirming that glucose was depleted after 13 h, inducing the consumption of ethanol (see supplementary material, Figure S4.2).

During this fermentation, the CGQ was found to be a very valuable tool for the monitoring of biomass, especially over night, when no samples for reference analysis were taken. Small variation in the sensor signal were not found to be a major disturbance and might even be valuable in terms of fermentation tracking with respect to operational events and troubleshooting.

4.3.2.2 Fermentation 2

A standard yeast fed-batch fermentation was performed as before with the difference that a constant particle concentration (potato protein powder) of 50 g/l was present during the fermentation. Glucose feeding of a 500 g/l stock solution was performed via a peristaltic pump (Reglo ICC, Ismatec, USA) adding the respective volume with a flow rate of 10 ml/min. The fermentation was followed by the CGQ for online biomass detection and by the NxPAS gas analyzer system for online off-gas monitoring (CO₂). The CO₂ profile can be considered as a replacement for OD data regarding the signal trend during the exponential growth phase. Accidentally, the CGQ was first started after 2 hours of the fermentation time. The course of the fermentation process is demonstrated in [Figure 4.11](#). Several processes events were induced as indicated.

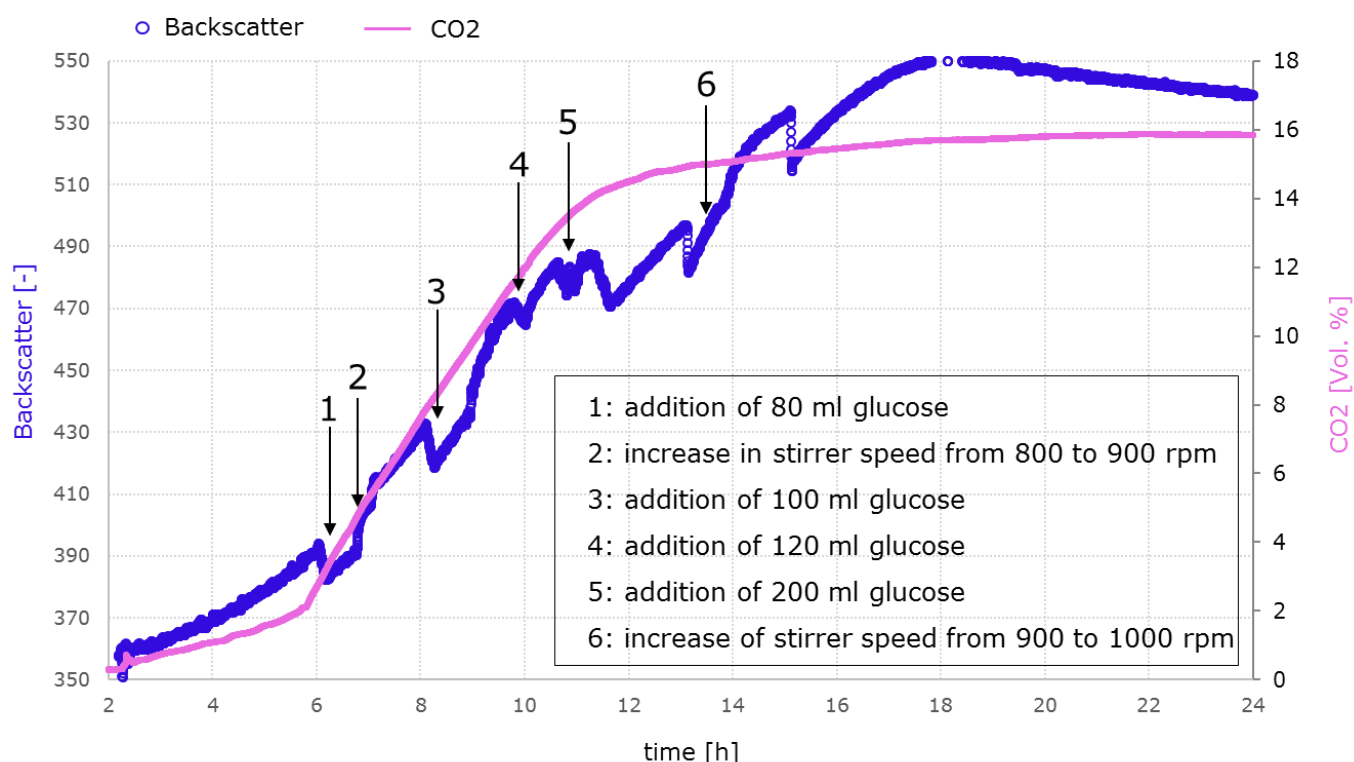


Figure 4.11: Biomass increase during a 24 h yeast fed-batch fermentation followed continuously on-line by the CGQ as backscatter intensity. The CGQ data was supplemented with on-line CO₂ off-gas data. Active operational events are indicated by black arrows and described in the legend. The signal drop after 13 h and 15 h was not related to an active process operation. It might be related to a drop in air pressure which sometimes happens during the fermentation due to a pressure drop in the supply line. It might also be related to a change in morphology, due to the diauxic shift.

Generally, the biomass increase during the fermentation was described by backscatter values between 350 and 550. The initial backscatter intensity of 350 is lower than the backscatter value obtained from the pre-test (see 4.3.1.3 YP medium containing particles). It can again be explained by dilution of the YP medium (and particle concentration) due to glucose addition. As before, feeding resulted in a small signal drop, explained by a strong local dilution in the detection area during feeding. The increase in the stirrer speed from 800 rpm to 900 rpm after approximately 7 hours resulted in a small, sharp increase in sensor signal. Contrarily, the increase in stirrer speed from 900 rpm to 1000 rpm after 13.5 h was hardly visible. It is assumed that the backscatter of light resulting from solid particles is higher than the backscatter of light resulting from air bubbles. Hence, it is furthermore expected that, after 13.5 h, the total concentration of solid particles was so high, that changes in the backscatter intensity caused by a change in air bubble size and concentration was minor. At this point, the sensor seemed to be less sensitive ('blind') towards changes in air-bubble dispersion. The signal drop after 13 h and 15 h was not due to an active operation. It is interesting to notice that the signal drop after 13 h happened at the end of glucose depletion. A respective drop after glucose depletion could also be observed in fermentation 1, which was also not assigned to an active operation. Hence, the drop after glucose depletion might be the result of a change in morphology due to the shift in metabolism towards metabolite consumption. The growth on ethanol occurs at a significant lower growth rate, in this fermentation indicated by the CO₂ profile.

Unfortunately, no HPLC analysis was performed here to prove this assumption. The dependency of yeast size on the growth rate (mean size increases with growth rate) is described by Tyson *et. al.* [14]. Smaller particles produce lower scattering resulting in a decrease in backscatter intensity. However, the drop after 15 h is unlikely to be explained by another metabolic shift, as the ethanol consumption was only in the beginning. Instead, the signal drop might be linked to a pressure drop in the air supply, decreasing the aeration rate. The signal line obtained by the CGQ appeared noisy due to the active inclusion of a lot of process events. However, the signal could be adjusted by considering the change in volume due to feeding and the correlation coefficients between the backscatter, the stirrer speed or the aeration rate. The presence of particles did not hinder the detection of biomass increase, which must be considered as a major advantage compared to OD measurements.

4.4 Conclusions

The CGQ(BIOR) was demonstrated to be a valuable tool for online, non-invasive biomass detection applying backscatter technology. Its application was demonstrated during a standard and particle containing yeast fed-batch fermentation. The application is facilitated via the glass wall of the reactor or a glass window inside the bioreactor wall of approximately 2 cm x 1.5 cm. The CGQ was found to be fast and easy to install and could be applied as a plug and play device. It was rather insensitive to the early stages of growth (below an OD value of 7, see [Figure 4.10 A](#)). However, in both fermentations, the entire process could be followed and no upper detection limit was reached within this study yielding a dry weight concentration of around 20 g/l and 30 g/l. The diauxic shift characterized by a significant decrease in growth rate after glucose depletion was clearly visible in the CGQ signal curve ([Figure 4.10 A](#)). It is anticipated that the shift is introduced by the sensor signal by a clear signal drop, resulting from a shrinking in cell size when adapting the metabolism towards metabolite consumption, coming along with a decrease in growth rate. Operational events were shown to affect the sensor signal in a mechanistic way. The effect of the aeration rate was found to be highly medium dependent. In the media under investigation (water, YP, YP containing 50 g/l particles) it was strongest in YP medium and lowest in YP medium containing particles, hence suggesting that the presence of solid particles make the sensor less sensitive towards changes in the air dispersion. The increase in backscatter intensity with increasing stirrer speed and aeration rate could be linked to an increase in air bubble concentration in the detection window, resulting from a better air dispersion by the stirrer at higher stirring rates. Pulse-feeding resulted in a drop in the sensor signal, due to a local dilution in the detection area.

The sensor can be considered as an easy-to-use and reliable tool for online biomass detection. Correlation to OD and dry weight yielded correlation coefficients of greater than 0.98 ([Figure 4.10 B](#) and [4.10 C](#)). The sensor signal was affected by process operations (as changes in stirrer speed, aeration rate and feeding) but this might be beneficial in terms of process tracking and

troubleshooting. A change in sensor signal due to operational events could most likely be circumvented by using correlations to correct the signal online.

4.5 Supplementary Material

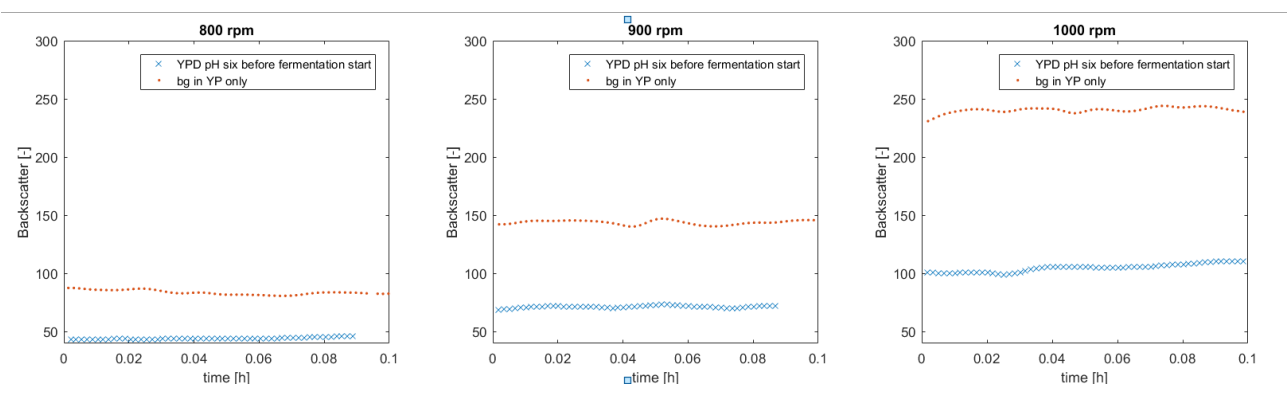


Figure S4.1: CGQ signal for the fermentation medium (fermentation 1) before (red dots) and after addition of 100 ml of a 400 g/l glucose stock solution (blue crosses). The addition of glucose diluted the YP medium, hence reducing the sensor signal.

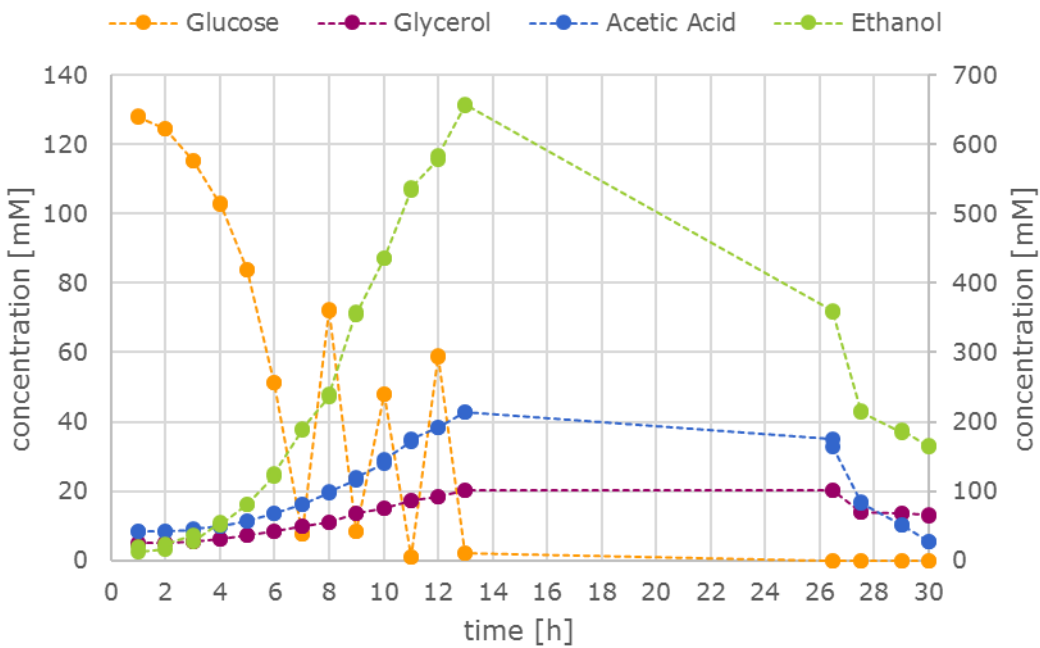


Figure S4.2: Glucose and metabolite profile for fermentation 1, detected by HPLC. The left y-axis refers to the concentration of glucose, glycerol and acetic acid. The right y-axis refers to the concentration of ethanol. Glucose feeding (100 ml of a 500 g/l glucose stock solution) was applied after 7 h, 9 h, and 11 h, resulting in a spike of glucose, each.

4.6 References

- [1] A. . Salgado, R. O. . Folly, and B. Valdman, "Biomass monitoring by use of a continuous on-line optical sensor," *Sensors Actuators B Chem.*, vol. 75, no. 1–2, pp. 24–28, Apr. 2001.
- [2] K. Kiviharju, K. Salonen, U. Moilanen, and T. Eerikäinen, "Biomass measurement online: the performance of in situ measurements and software sensors," *J. Ind. Microbiol. Biotechnol.*, vol. 35, no. 7, pp. 657–665, Jul. 2008.
- [3] V. Camisard, J. P. Brienne, H. Baussart, J. Hammann, and H. Suhr, "Inline characterization of cell concentration and cell volume in agitated bioreactors using in situ microscopy: Application to volume variation induced by osmotic stress," *Biotechnol. Bioeng.*, 2002.
- [4] S. Beutel and S. Henkel, "In situ sensor techniques in modern bioprocess monitoring," *Appl. Microbiol. Biotechnol.*, vol. 91, no. 6, pp. 1493–1505, Sep. 2011.
- [5] C. Slouka *et al.*, "Low-Frequency Electrochemical Impedance Spectroscopy as a Monitoring Tool for Yeast Growth in Industrial Brewing Processes," *Chemosensors*, vol. 5, no. 3, p. 24, Aug. 2017.
- [6] R. E. Madrid and C. J. Felice, "Microbial Biomass Estimation," *Crit. Rev. Biotechnol.*, vol. 25, no. 3, pp. 97–112, Jan. 2005.
- [7] D. Watson, N. Hagen, J. Diver, P. Marchand, and M. Chachisvilis, "Elastic Light Scattering from Single Cells: Orientational Dynamics in Optical Trap," *Biophys. J.*, vol. 87, no. 2, pp. 1298–1306, Aug. 2004.
- [8] S. Bruder, M. Reifenrath, T. Thomik, E. Boles, and K. Herzog, "Parallelised online biomass monitoring in shake flasks enables efficient strain and carbon source dependent growth characterisation of *Saccharomyces cerevisiae*," *Microb. Cell Fact.*, vol. 15, no. 1, pp. 1–14, 2016.
- [9] L. Øgendal, "Light Scattering a brief introduction," *Univ. Copenhagen*, no. February, pp. 3–4, 2016.
- [10] N. G. Khlebtsov, "Introduction to Light Scattering by Biological Objects. Part 1. Extinction and Scattering of Light in Disperse Systems," 2002.
- [11] J. Schmidt-Hager, C. Ude, M. Findeis, G. T. John, T. Scheper, and S. Beutel, "Noninvasive online biomass detector system for cultivation in shake flasks," *Eng. Life Sci.*, vol. 14, no. 5, pp. 467–476, 2014.
- [12] A. Dunn and R. Richards-Kortum, "Three-dimensional computation of light scattering from cells," *IEEE J. Sel. Top. Quantum Electron.*, vol. 2, no. 4, pp. 898–905, 1996.
- [13] O. B. Monitoring, "CGQ BioR."
- [14] C. B. Tyson, P. G. Lord, and A. E. Wheals, "Dependency of Size of *Saccharomyces cerevisiae* Cells on Growth Rate," *J. Bacteriol.*, vol. 138, no. 1, pp. 92–98, 1979.

Final Conclusions

Within this thesis, 4 different technologies have been studied with respect to their application as a monitoring tool for fermentation processes. The technologies under study, exemplarily applied to a lab scale yeast fermentation process, comprised

- 1) the detection of glucose by means of an electrochemical biosensor,
- 2) the quantification of the 6 analytes glucose, ethanol, glycerol, acetic acid, ammonium and phosphate by means of IR spectroscopy and chemometric modelling,
- 3) the detection of microbial growth complemented with morphological studies over the fermentation time by means of imaging and image analysis, and
- 4) the detection of microbial growth by means of a backscatter-sensor, measuring the increase of biomass through the fermenter wall in case of a glass vessel or through a small glass window, respectively.

All technologies can be considered for on-line process data collection.

The biosensor, designed as a small flow through cell, facilitated fast off-line and continuous on-line detection of glucose during a 10 h fermentation process. Glucose detection was facilitated in a concentration range from 1mM to 150 mM, applying a segmented calibration curve for low (1 mM – 60 mM) and high (60 mM – 150 mM) glucose concentrations. By means of the biosensor, glucose concentrations are transferred into an electric current via the enzyme glucose oxidase, requiring a minimum oxygen availability for its activity. Off-line, the measurement was compromised by oxygen limitations when both, high glucose and high cell concentrations were present in the fermentation sample. However, when applied on-line and positioned closely to the fermenter outlet, this is expected to be circumvented as the sensor is supplied constantly with fresh, aerated broth. The microfluidic device was limited by a maximal flow -rate of 1 ml/min. I consider the sensor as especially valuable for the early stages of fermentation process development, when microbial kinetics are still unknown. Measurements with the sensor will significantly reduce the time spent on process optimization. Applying the sensor, a first insight into the glucose consumption can be gained right away, allowing for optimization with respect to glucose feeding 'on the first go'. With its design it appears particularly suitable for lab or pilot scale experiments. To which extent the sensor stays active during long-term processes running several days or even weeks, would be an interesting aspect to study in the future.

IR-spectroscopy combined with PLS-modelling was found to be a highly powerful tool for rapid multi-component quantification in the complex fermentation matrix. Within this work, glucose, ethanol, and to a lesser extent ammonium and phosphate were modeled based on distinct spectral features, demonstrating a prediction error of less than 15 %. The calibration of glycerol and acetic acid was challenged by low concentration levels within the process, relatively low IR activity compared to the other components, and cross-correlation of spectral absorption bands. Thus, PLS-calibration models for glycerol and acetic acid were shown to rely on indirect prediction, trapped in

a cave of covariance with ammonium and phosphate, or respectively, regions linked to the spectral absorbance of ammonium and phosphate. Hence, the relative RMSEP for ammonia and phosphate resulted in values between 50 % – 200 %. This study pointed out the importance of multivariate calibration design and the evaluation of the obtained PLS-models beyond the common calibration and validation parameters as R^2 , bias and RMSE. Models were built based on spectra acquisition obtained by the novel NLIR technology. The particular application range for the NLIR instrument is designed for the region 900 cm^{-1} – 1550 cm^{-1} , which is rich in spectral information related to the functional groups comprising most bio-molecules. However, the calibration process is complex and intense and might only pay off when having a long-term view in mind and providing consistent process conditions. This applies to industrial production where bioprocesses must be operated under optimized and defined conditions. In such an environment, IR-spectroscopy combined with PLS modelling is expected to be a powerful tool, ensuring consistent processing and quality of biotechnological products. So far, MIR has not found entry to industrial processing as a monitoring tool, due to its high cost and fragile moving part in common FT-IR spectrometers. However, these issues are eliminated by novel technologies as demonstrated by the NLIR instrument, tailored towards industrial needs.

When starting this work, the oCelloScope instrument, facilitating automated imaging and image analysis of the growing yeast culture, was used off-line and fermentation samples were studied via a 96 well-plate. The need for a flow-through cell was addressed by ParticeTech Aps facilitating a first continuous imaging set-up by means of a prototype flow-through-cell. For the very first time, the technology in use was connected continuously to the fermenter. Hence, imaging and image analysis could be performed automatically every 10 minutes during a 10 h yeast (fed-) batch fermentation. On-line images could be used in particular for the detection of yeast growth, while off-line image data of diluted fermentation samples, complementing the on-line data, could be used to yield insights into the morphology dynamics of the growing yeast culture.

OD values between 0.6 and 4 could be followed reliably on-line by means of the BCA algorithm, providing increased sensitivity compared to OD measurements. However, cell concentrations referring to an OD value exceeding 4 (a cell dry weight of approximately 2 g/L) were found outside the capacity of image based growth detection by the algorithms applied in this study. As the growth detection was limited by the absorbance by the sample, a reduction of the height of the flow-through-cell is expected to enable the continuous detection of microbial growth up to an OD value of approximately 22. The quality of the on-line images was deficient for morphology analysis due to the presence of too high cell concentrations and an out-of-focus issue progressing over time. The latter resulted from a bottom deformation of the prototype flow-through cell. However, first morphological trends could be obtained by the images acquired off-line on diluted fermentation samples. The early stages of (exponential) growth showed a slight shift towards cell clusters, the late stages of (exponential) growth showed a slight shift towards the occurrence of single cells. However, budding cells were the dominating cell species within this study, investigating the culture over an extended batch phase.

Providing appropriate image quality with respect to image analysis especially regarding morphological studies was found to be a crucial requirement. Challenges linked to the flow- through - cell and the resulting image quality were addressed by a mold for improved flow-through-cell fabrication as well as a dual pump flow controller for automatic sample dilution.

Thus, by now, a promising and exciting set-up is ready to continue the research based on image analysis. Findings in this thesis might be used to transform the knowledge gained into a novel, image based monitoring strategy, enabling growth detection and moreover the assessment of crucial fermentation parameters based on cell morphology.

The CGQ(BIOR) was demonstrated to be a valuable tool for online, non-invasive biomass detection applying backscatter technology. The application is facilitated via the glass wall of the reactor or a glass window inside the bioreactor wall, that must have a minimum size of approximately 2 cm x 1.5 cm. It was found to be especially easy to use, it could be (de-) installed any time during the fermentation process and no sterilization or intense cleaning was required. It was possible to follow the microbial growth during the entire fermentation process under study yielding a final OD value of 55 and a respective cell dry weight concentration of 20 g/l. The sensor was rather insensitive to the early stages of growth (below an OD value of 7). However, the microbial growth during the entire process could be followed, even in the presence of a constant particle concentration of 50 g/l. No upper detection limit was observed within this study. Correlation to OD and dry weight yielded correlation coefficients greater than 0.98, which is an excellent achievement. The sensor signal was affected by process operation conditions (as changes in stirrer speed, aeration rate and feeding) but this might be beneficial in terms of process tracking and troubleshooting. The change in the sensor signal due to a change in stirrer speed and aeration rate could be explained by a change in the air bubble size and concentration in the detection window, resulting from different stirring rates (aeration rates). It was interesting to observe a drop in the sensor signal, directly after the diauxic shift, during the two fermentations performed. It is anticipated that this drop might result from the significant decrease in growth rate, described to come along with a decrease in the yeast cell size. Smaller particles (cells), in turn, yield lower backscattering, hence reducing the sensor signal. It would be very interesting to combine the CGQ and the oCelloScope instrument, in order to investigate the effect observed. I consider the CGQ as a next generation OD measurement. The sensor signal is certainly limited at high cell concentrations, and the limit is to be investigated. However, the sensor yielded representative growth data, with practically no effort.

It must be realized that especially the biosensor and the backscatter technology under study, both commercially available and ready to use, facilitated the monitoring of two crucial process parameters, glucose concentration and microbial growth, practically with minimum effort. It evidently demonstrates that technology for the monitoring of crucial fermentation parameters is available, and already tremendously reduces the labor intensity in the laboratory. However, trust and knowledge on available technology can only be gained and deepened when considered and used on a daily basis. Even more advanced tools might be available in the future. Nonetheless, I think that convention and standards need to change gradually, applying the possibilities that are available right now. Otherwise given opportunity might be lost on the way. By committing the present technical progress, we can only benefit in terms of both, knowledge and understanding of the process and on the instrumentation, pushing towards a generic goal. This also holds for equipment under development as the NLIR instrument and the oCelloScope equipment.

Appendix 1

Manuscript Details

Manuscript number	NBT_2019_41
Title	Fast measurement of phosphates and ammonium in fermentation-like media: feasibility study
Article type	Full Length Article

Abstract

Real-time monitoring of bioprocesses plays a key-role in modern industries, providing new knowledge on full-scale production thus enabling control on the process, allowing to run it at optimal conditions while minimizing waste. The monitoring of phosphates and ammonium in fermentation processes has a double interest: they are important nutrients of any living organism and at the same time, from an environmental point of view, constitute nutrient pollutants and their unnecessary use and disposal must be avoided. In this work, we verified the possibility of simultaneous analysis of phosphates and ammonium in fermentations, using spectroscopy-based methods combined with chemometrics to construct calibration models. To achieve this, we built the calibration models on synthetic samples mimicking real fermentation media providing a dataset where the analytes were completely uncorrelated. Different at-line techniques (Mid- and Near- infrared spectroscopy, MIR and NIR) were evaluated for their ability to quickly monitor both analytes, in a wide range of concentration (10-100 mM), in three media of different complexity. Partial Least Squares (PLS) models on MIR spectroscopy gave good results, with prediction errors lower than 5% for both analytes, in all datasets. PLS models on NIR spectroscopy provided, instead, inferior results (prediction errors between 3 and 26%) for both analytes, and a model built on indirect correlations for phosphates.

Keywords	Process Monitoring; Near-infrared spectroscopy; Mid-infrared spectroscopy; Chemometrics; Fermentation.
Corresponding Author	Marta Bevilacqua
Corresponding Author's Institution	Copenhagen University
Order of Authors	Katrin Pontius, Giulia Praticò, Flemming H. Larsen, Thomas Skov, Nils Arneborg, Anna Eliasson Lantz, Marta Bevilacqua
Suggested reviewers	Anurag Rathore, Jarka Glassey, Edy Brito, Sylvie Roussel

Submission Files Included in this PDF

File Name [File Type]

CoverLetter.docx [Cover Letter]

Highlights_FermentationParameter.docx [Highlights]

Paper_FermentationParameter.docx [Manuscript File]

Fig1.tif [Figure]

Fig2.tif [Figure]

Fig3.tif [Figure]

Fig4.tif [Figure]

Fig5.tif [Figure]

Fig6.tif [Figure]

FigA1.tif [Figure]

FigB1.tif [Figure]

FigB2.tif [Figure]

Author declaration signed.pdf [Author Agreement]

SupplementaryMaterial.pdf [e-Component]

To view all the submission files, including those not included in the PDF, click on the manuscript title on your EVISE Homepage, then click 'Download zip file'.

Dr. Mike Taussig
Editor-in-Chief
New Biotechnology

February 01, 2019

Dear Dr. Mike Taussig,

I am pleased to submit the manuscript entitled "Fast measurement of phosphates and ammonium in fermentation-like media: feasibility study" by Katrin Pontius, Giulia Praticò, Flemming H. Larsen, Thomas Skov, Nils Arneborg, Anna Eliasson Lantz and Marta Bevilacqua. We kindly propose its consideration as an original research article for publication in "New Biotechnology".

In this manuscript, we compared the potential of two vibrational spectroscopic techniques (Mid- and Near-Infrared) for fast quantification of phosphates and ammonium in fermentation processes. We decided to build chemometric calibrations models on synthetic samples mimicking real fermentation media. Thus, we ensured the absence of correlations between these two analytes in the samples under study. Decoupling the correlations between the analytes under study is a crucial aspect with respect to building chemometric calibration models. Our method was proven successfully (with prediction errors lower than 5%) in the simultaneous quantification of both compounds and in a wide range of concentration (10-100 mM).

We believe that the approach proposed could be easily adapted for fast monitoring and optimization in a large variety of bioprocesses addressing the need of suitable monitoring strategies in the fermentation era. Therefore, we envision that this is of significant interest for the readers of "New Biotechnology".

This manuscript has not been published and is currently not under consideration for publication elsewhere. All authors have read the current version of the manuscript and agreed on its submission to "New Biotechnology". None of the authors has any conflict of interest to disclose.

Thank you for your consideration.

Yours sincerely,

Dr. Marta Bevilacqua

Department of Food Science, University of Copenhagen
Rolighedsvej 26
DK-1958 Frederiksberg C, Denmark
marta@food.ku.dk

1 HIGHLIGHTS

- 2 - Real-time monitoring of fermentation parameters is important for modern industries.
- 3 - Analysis of phosphates and ammonium in a wide range of concentration (10-100 mM).
- 4 - Use of synthetic samples provided a dataset with completely uncorrelated analytes.
- 5 - Mid and Near infrared spectroscopy were evaluated in fermentation-like media.
- 6 - Chemometric models on Mid-infrared spectra had prediction errors lower than 5%.

7

Fast measurement of phosphates and ammonium in fermentation-like media: feasibility study

Katrin Pontius^a, Giulia Praticò^{b,c}, Flemming H. Larsen^b, Thomas Skov^b, Nils Arneborg^b, Anna Eliasson Lantz^a, Marta Bevilacqua^{*b}

^a Department of Chemical and Biochemical Engineering, Technical University of Denmark, Søtofts Plads, Building 229, 2800 Lyngby

^b Department of Food Science, University of Copenhagen, Rolighedsvej 26, DK-1958 Frederiksberg C, Denmark.

^c Department of Nutrition, Exercise and Sports, University of Copenhagen, Rolighedsvej 26, DK-1958 Frederiksberg C, Denmark.

Declarations of interest: none

*** CORRESPONDING AUTHOR:** Marta Bevilacqua, Department of Food Science, University of Copenhagen, Rolighedsvej 26, DK-1958 Frederiksberg C, Denmark.

e-mail: marta@food.ku.dk phone number: +45 52729860

ORCID: 0000-0002-3361-9219

1 **ABSTRACT**

2 Real-time monitoring of bioprocesses plays a key-role in modern industries, providing new
3 knowledge on full-scale production thus enabling control on the process, allowing to run it at
4 optimal conditions while minimizing waste. The monitoring of phosphates and ammonium in
5 fermentation processes has a double interest: they are important nutrients of any living organism
6 and at the same time, from an environmental point of view, constitute nutrient pollutants and their
7 unnecessary use and disposal must be avoided.

8 In this work, we verified the possibility of simultaneous analysis of phosphates and ammonium in
9 fermentations, using spectroscopy-based methods combined with chemometrics to construct
10 calibration models. To achieve this, we built the calibration models on synthetic samples mimicking
11 real fermentation media providing a dataset where the analytes were completely uncorrelated.

12 Different at-line techniques (Mid- and Near- infrared spectroscopy, MIR and NIR) were evaluated
13 for their ability to quickly monitor both analytes, in a wide range of concentration (10-100 mM), in
14 three media of different complexity. Partial Least Squares (PLS) models on MIR spectroscopy gave
15 good results, with prediction errors lower than 5% for both analytes, in all datasets. PLS models on
16 NIR spectroscopy provided, instead, inferior results (prediction errors between 3 and 26%) for both
17 analytes, and a model built on indirect correlations for phosphates.

18

19 **KEY-WORDS**

20 Process Monitoring; Near-infrared spectroscopy; Mid-infrared spectroscopy; Chemometrics;
21 Fermentation.

22

1 ABBREVIATIONS

2 ATR: attenuated total reflectance; CV: Cross-Validation; d1: recycle delay; DTGS: deuterated
3 triglycine sulfate detector; FT: Fourier Transformed; h_i : leverage of the i th validation sample; i :
4 samples index; MIR: mid-infrared spectroscopy; MSC: Multiplicative scatter Correction; MSEC .
5 mean squared error in calibration; N : number of samples; NET: net analyte signal; NIR: mid-
6 infrared spectroscopy; NMR: Nuclear Magnetic Resonance spectroscopy; o.d.: Optical Denisty; P_i :
7 inorganic phosphates; PLS: Partial Least Squares regression; R^2 : coefficient of determination;
8 RMSE: Root Mean Squared Error; SiC: Global silicon Carbide; SNV: Standard Normal Variate;
9 $s_{PE_i}^2$: prediction uncertainty (variance); s_{PE_i} : standard deviation; UV-VIS: Ultra-Violet Visible
10 spectroscopy; VIP: Variable Importance in Prediction.

11

12

13

1 INTRODUCTION

2 Due to the constant need for more efficient and sustainable biotech production achieved by reducing
3 the consumption of energy and raw materials while improving yields, process understanding is
4 particularly important. In this context, real-time monitoring of bioprocesses plays a dual key-role.
5 On one side, it allows industries to gain new knowledge on full-scale production and, hence, to be
6 more prudent in the use of resources. On the other side, it also enables control on the process,
7 giving the possibility to run it at more optimal conditions, and thereby minimizing waste. Both
8 aspects hold not only for industrial production but are equally important for efficient fermentation
9 optimization in prior steps.

10 In this framework, the monitoring of phosphates and ammonium in fermentation processes is of
11 particular interest. Phosphates and ammonium are important nutrients of any living organism and,
12 as such, their monitoring is crucial to avoid a deficiency in the fermentation broth. At the same
13 time, from an environmental point of view, they constitute nutrient pollutant and their unrestricted
14 disposal in the environment affects the water resources worldwide, compromising ecosystems and
15 human health. Finally, phosphorus is present in a finite amount on the planet and its unnecessary
16 use and disposal must be carefully considered by any industry [1, 2]. Consequently, for sustainable
17 and economically viable production, excess addition should be avoided.

18 One of the most popular methods to quantify various analytes in fermentation broth is by means of
19 colorimetric methods. These are based on chemical or enzymatic reactions that selectively act on
20 the compounds to be measured transforming them into different species that can be observed and
21 quantified by UV-VIS spectrophotometry. This kind of analysis is generally applicable only in
22 limited concentration ranges and involves a pretreatment of the sample before analysis: samples
23 must therefore be taken from the bioreactor, filtered, diluted, and conditioned before analysis [3].

1 The main drawback of this approach is the time delay in the process data between the sampling and
2 the result of the analysis. On average, colorimetric assays take at least 30 min of time before a result
3 is available. Note that, samples taken during the process are usually collected and analyzed after the
4 process has finished due to practical reasons as equipment scheduling and operational shifts. Thus,
5 results obtained from the assays stating the fermentation progress and performance are commonly
6 available only after the overall fermentation has finished. In contrast, real-time techniques have the
7 advantage of providing a direct insight into the bioprocess that is essential, in case of problems, for
8 their early detection, allowing an immediate action to resolve the situation without stopping the
9 process [4, 5]. Nevertheless, real-time methods are still quite scarcely applied to industrial scale
10 bioprocesses due to practical difficulties: on one side, the analysis must preserve the sterility of the
11 process, while on the other side, it must be able to measure specific analytes, sometimes at very low
12 concentration, in presence of complex nutritive media used for the process [5]. In the past years,
13 different strategies have been proposed for on-line monitoring of phosphates and ammonium during
14 fermentation and other biological processes [3, 6-10]. Among these, the most promising techniques
15 implemented in on-line or at-line set up are vibrational spectroscopies (namely near-infrared and
16 mid-infrared spectroscopy), coupled with mathematical modelling that can relate the spectra
17 obtained to the important process parameters to be measured [4, 11-18]. Spectroscopic techniques
18 have the advantage of enabling rapid analysis (< 3 min per spectrum depending on the instrument
19 specifications and settings, in the fastest case only about 10-20 seconds) and they can be applied on-
20 line to the bioreactors by means of fiber-optic probes. Besides, multiple analytes can be quantified
21 based on a single spectrum. When analyzing fermentation samples, though, mathematical models
22 are generally challenged by the presence of many different components in the broth, all absorbing in
23 the same spectral regions and quite often being directly or indirectly correlated to each other.
24 Consequently, the resulting spectra have a high number of broadly overlapping spectral features

changing throughout the course of the fermentation. Multivariate calibration models for such processes will therefore often be valid only for modeling a process conducted under the same conditions as the one used for building it, and they will fail if any slight unforeseen change occurs in the future. In order to have meaningful and robust calibration models the analyte of interest in the calibration samples should not be correlated with any other species present in the medium [19]. This is rarely true in real-life fermentation processes. Therefore, there is a need of decoupling (or uncorrelating) the concentration dynamics of the target analytes from the ones of other compounds present, introducing some artificial, uncorrelated variations but keeping the matrix of the samples as representative as possible of the original fermentation medium.

Although some authors argue that it is impossible to ignore the correlations/ interactions occurring among media components [15, 20-22], others advocate for the need of overcoming the natural correlations and have addressed this challenge in different ways that mainly fall into two approaches [11, 16-18, 23-28]. The first is the preparation of synthetic samples (alone or in combination with the presence of real samples in the model) [16-18, 23] while the second is an in-between situation where samples are spiked with different analytes to break the correlations (so called, semi-synthetic samples or adapted samples) [24-28].

We decided to verify the possibility of simultaneous analysis of phosphates and ammonium in fermentation samples using spectroscopy-based methods and building the calibration on synthetic samples. Synthetic samples were generated aiming at resembling real fermentation media and at the same time providing a dataset where the analytes were completely uncorrelated. Although the determination of ammonia by both MIR (mid-infrared) and NIR (near-infrared) spectroscopy has been presented in the literature before [4, 11], the research of a technique able to simultaneously quantify phosphates and ammonium in fermentation broth and in such wide ranges of concentrations (10-100 mM) has not been addressed so far.

In this study, NIR and MIR spectroscopy were evaluated and compared regarding their ability to determine the concentration of the analytes under study. NIR spectroscopy has the advantage of being the easiest to implement on-line on an industrial-scale fermenter, whereas MIR spectroscopy has the possibility to be implemented on-line as well, but it requires shorter optical fibers and therefore a closer position to the reactor. The mid range of IR is generally more sensitive, informative, respectively, therefore providing a better chance for simultaneous detection of the two analytes in the lower concentration range.

MATERIALS AND METHODS

PREPARATION OF THE SAMPLES

All sets of samples were prepared using precision lab glassware accurately washed and then soaked overnight in a solution of 5% (v/v) HNO_3 prepared in distilled water. The labware was rinsed at least three times in distilled water and subsequently dried before use.

To efficiently study the possibility of determination of phosphates and ammonia by fast spectroscopic techniques, three different sets of synthetic samples have been prepared. For all three sets, 25 solutions have been prepared containing both analytes in concentrations ranging from 10 – 100 mM equally spaced according to a full factorial design (Figure 1 and Table S1). The solutions of different concentrations were obtained by preparing, per each set of samples, two bulk solutions of the two salts (0.4M for Na_2HPO_4 , and 4M for NH_4Cl) in pure milliQ water, minimal medium and complex medium, respectively. The chemicals used were: $\text{Na}_2\text{HPO}_4 \cdot 2\text{H}_2\text{O}$ (Merck KGaA, Darmstadt, Germany; lot 1.06580.1000) and NH_4Cl (Sigma-Aldrich Co., St. Luis, MO, USA; lot SZBD1910V). Then, proper volumes of these two solutions were mixed in order to reach the desired concentration in a final volume of 50 ml. This final volume was then adjusted by using milliQ water, minimal medium and complex medium, respectively. The pH in all the bulk solutions

was adjusted to 6.0 ± 0.1 using a few drops of 6M HCl (using Edge pH-meter with digital pH electrode, Hanna Instruments Inc., Woonsocket, RI, USA). The pH of 6.0 was chosen for all the sets of samples to avoid the precipitation of insoluble salts that occurred during some initial tests of minimal medium preparation. Investigating the precipitation observed in the medium with the free-ware Visual MINTEQ 3.1 [29], we found it to be due to the formation of struvite ($\text{NH}_4\text{MgPO}_4 \cdot 6\text{H}_2\text{O}$) and hydroxyapatite ($\text{Ca}_5(\text{PO}_4)_3(\text{OH})$) [30, 31]. Using the same software, the theoretical precipitation at pH 6.0 was calculated as negligible, and thus this pH was applied in this work. Besides, pH 6 is a widely used level for pH control as optimal for growth and production strategies in fermentation processes.

The minimal medium used for the second set of samples contained per liter of milliQ water: 9 g $\text{MgSO}_4 \times 7 \text{H}_2\text{O}$ (Sigma, Bioreagent for molecular biology, $\geq 99.0\%$ lot SLBR2370V) 0.42 g $\text{CaCl}_2 \times 2\text{H}_2\text{O}$ (Sigma-Aldrich, ACS reagent, $\geq 99.0\%$, lot BCBK4048V) and 13.2 g Glucose (Alfa Aesar, anhydrous, 99 %). 40 ml of trace metal solution and 40 ml of vitamin solution were added to 1L of medium based on the recipe as described in Christiansen et al. [32].

The complex media used in the third set of samples was prepared by dissolving per liter of milliQ water: 9 g $\text{MgSO}_4 \times 7 \text{H}_2\text{O}$ (Bioreagent for molecular biology, $\geq 99.0\%$ lot SLBR2370V, Sigma) and 0.42 g $\text{CaCl}_2 \times 2\text{H}_2\text{O}$ (Sigma-Aldrich, ACS reagent, $\geq 99.0\%$, lot BCBK4048V). The solution was heated to 55°C and the pH was adjusted to 6.5-7.5 using a few drops of 2 M citric acid. The complex source (potato protein, 55 g/l) was added and enzymatically hydrolyzed using 2.75 g Alcalase (Alcalase® 2.4 L FG (EC 3.4.21.62), Unit Activity 2.4 AU-A/g, 105 Novozymes, Bagsvaerd, Denmark) for 2 h at 55°C . Subsequently, the mixture was centrifuged (10 min, 4700 rpm) and the supernatant filtered through a $20 \mu\text{m}$ filter (filter papers 41, CAT No. 1441-055, Whatman) before being autoclaved. After autoclavation, 40 ml of trace metal solution and 40 ml of vitamin solution according to the recipe as described in Christiansen et al. [32] were added to 1L of

was adjusted to 6.0 ± 0.1 using a few drops of 6M HCl (using Edge pH-meter with digital pH electrode, Hanna Instruments Inc., Woonsocket, RI, USA). The pH of 6.0 was chosen for all the sets of samples to avoid the precipitation of insoluble salts that occurred during some initial tests of minimal medium preparation. Investigating the precipitation observed in the medium with the free-ware Visual MINTEQ 3.1 [29], we found it to be due to the formation of struvite ($\text{NH}_4\text{MgPO}_4 \cdot 6\text{H}_2\text{O}$) and hydroxyapatite ($\text{Ca}_5(\text{PO}_4)_3(\text{OH})$) [30, 31]. Using the same software, the theoretical precipitation at pH 6.0 was calculated as negligible, and thus this pH was applied in this work. Besides, pH 6 is a widely used level for pH control as optimal for growth and production strategies in fermentation processes.

The minimal medium used for the second set of samples contained per liter of milliQ water: 9 g $\text{MgSO}_4 \times 7 \text{H}_2\text{O}$ (Sigma, Bioreagent for molecular biology, $\geq 99.0\%$ lot SLBR2370V) 0.42 g $\text{CaCl}_2 \times 2\text{H}_2\text{O}$ (Sigma-Aldrich, ACS reagent, $\geq 99.0\%$, lot BCBK4048V) and 13.2 g Glucose (Alfa Aesar, anhydrous, 99 %). 40 ml of trace metal solution and 40 ml of vitamin solution were added to 1L of medium based on the recipe as described in Christiansen et al. [32].

The complex media used in the third set of samples was prepared by dissolving per liter of milliQ water: 9 g $\text{MgSO}_4 \times 7 \text{H}_2\text{O}$ (Bioreagent for molecular biology, $\geq 99.0\%$ lot SLBR2370V, Sigma) and 0.42 g $\text{CaCl}_2 \times 2\text{H}_2\text{O}$ (Sigma-Aldrich, ACS reagent, $\geq 99.0\%$, lot BCBK4048V). The solution was heated to 55°C and the pH was adjusted to 6.5-7.5 using a few drops of 2 M citric acid. The complex source (potato protein, 55 g/l) was added and enzymatically hydrolyzed using 2.75 g Alcalase (Alcalase® 2.4 L FG (EC 3.4.21.62), Unit Activity 2.4 AU-A/g, 105 Novozymes, Bagsvaerd, Denmark) for 2 h at 55°C . Subsequently, the mixture was centrifuged (10 min, 4700 rpm) and the supernatant filtered through a 20 μm filter (filter papers 41, CAT No. 1441-055, Whatman) before being autoclaved. After autoclavation, 40 ml of trace metal solution and 40 ml of vitamin solution according to the recipe as described in Christiansen et al. [32] were added to 1L of

1 medium. 100 ml of 132 g/l Glucose solution and 1 ml of aqueous-silicon based Antifoam (Antifoam
2 Y-30 Emulsion, Sigma-Aldrich) were added finally to reproduce a true complex fermentation
3 medium.

4
5 **Figure 1. DoE representation.**

6
7 **REFERENCE MEASUREMENTS**

8 **Phosphate Colorimetric Assay kit.** Phosphate concentrations were determined in duplicates after
9 diluting to a proper concentration using the Phosphate Colorimetric Assay KIT from Sigma Aldrich
10 (MAK030). During this assay inorganic phosphates (P_i) reacts with a chromogenic complex, which
11 results in a colorimetric (650 nm) product proportional to the amount of phosphates present.
12 Samples were analyzed on a 96 well plate (Costar Assay Plates, REF. 3370) with the
13 spectrophotometer Multiskan Go from Thermo Scientific.

14 **Ammonium Colorimetric Assay kit.** Ammonium concentrations were determined in duplicates
15 after diluting to a proper concentration using the Ammonium Cuvette -Test 47-130 mg/L NH_4-N
16 (LCK 302) from Hach Lange. In this assay ammonium ions react at pH 12.6 with hypochlorite and
17 salicylate ions in the presence of sodium nitroprusside as a catalyst to form indophenol blue
18 (measured at 694 nm). Samples were analyzed in the Hach Lange Spectrophotometer DR 3900.

19 **^{31}P NMR spectroscopy.** For the experiments in water and minimal media, samples for ^{31}P NMR
20 spectroscopy were prepared in 5 mm (o.d.) NMR sample tubes by mixing 495 μ l of sample with 55
21 μ l of D_2O . For the samples from the complex media and the fermentation, 440 μ l of samples were
22 mixed with 110 μ l of D_2O . All samples were analyzed using a Bruker Avance DRX 500 (11.7 T)

1 spectrometer operating at a Larmor frequency of 202.4 MHz for ^{31}P , using a 90 degrees single pulse
2 experiment. For each experiment, 16k data points were acquired with 256 (water and minimal
3 medium) or 512 scans (complex medium), with a spectral width of 8 kHz and an acquisition time of
4 1.01 s. Recycle delay (d1) was properly optimized for each sample matrix in order to obtain a
5 quantitative recovery of the signal. Particularly a d1 of 64 s was used for samples in water and a d1
6 of 2 s was used for the rest of the samples. Receiver gain was kept fixed at a value of 9195. For the
7 absolute quantification of phosphorus, an external reference of 86.73 mM phosphoric acid (H_3PO_4)
8 was prepared from a solution of H_3PO_4 85 wt. % (AppliChem GmbH, Darmstadt, Germany) in two
9 sets of two replicates each, containing 10% and 20% of D_2O respectively, in order to reflect the
10 concentration of the samples analyzed. Reference samples were acquired with the same settings
11 reported for the previous samples, with a d1 of 64 s. The absolute phosphates concentration of each
12 sample was obtained comparing the integral of each phosphates peak with the averaged integral of
13 the two replicates of the external standard of known concentration. The free induction decay was
14 apodized by exponential line broadening of 1 Hz prior to Fourier transformation. Each spectrum
15 was manually phased and baseline corrected. Spectra were processed by using Bruker TopSpin 3.5
16 (Bruker BioSpin GmbH, Germany) and data were imported in MatLab 8.3.0.532 (MathWorks,
17 Natick, MA).

18 SPECTROSCOPIC MEASUREMENTS

19 **Infrared spectroscopy (IR).** Infrared spectra of all the solutions were acquired on a FT-
20 spectrometer (ABB Bomem MB100) equipped with a Globar silicon Carbide (SiC) radiation source
21 collimated to a 2.5 cm diameter beam and a deuterated triglycine sulfate (DTGS) detector. The
22 measurements were performed in attenuated total reflectance (ATR) mode, by putting 800 μL of
23 each sample on the ATR cell (ZnSe crystal, 45°C) and recording a full spectrum between 750 and

1 4000 cm^{-1} with 64 scans at a resolution of 4 cm^{-1} ¹. The background was obtained on an empty ATR
2 groove using 128 scans. Each sample was acquired in duplicate. In between samples the ATR
3 groove was cleaned with water and dried out with spectrophotometric tissues. Subsequently only
4 the non-noisy region, between 800 and 2250 cm^{-1} , was kept for further analysis. PLS models were
5 built on the whole spectral region (800 – 2250 cm^{-1}) as well as on the region related to the target
6 species (950 – 1200 cm^{-1} for phosphates and 1360 – 1540 cm^{-1} for ammonium).

7 **Near Infrared spectroscopy (NIR).** Near infrared spectra were acquired using a
8 dispersive/scanning-rating NIR instrument (Foss DS2500, Foss Analytical A/S, Hilleroed,
9 Denmark) equipped with a halogen lamp as the light source and a PbS detector. The measurements
10 were done with the slurry cup sample holder, using a golden reflector allowing an optical path of 1
11 mm (Foss DS2500, Foss Analytical A/S, Hilleroed, Denmark). Each measurement was carried out
12 by averaging 16 scans taken in 4 different position of the sample holder. For each sample, 3
13 different aliquots of about 500 μL each were analyzed in this way and a full spectrum, between 400
14 and 2500 nm, was recorded for each of them¹. In between samples the sample holder was cleaned
15 with water and ethanol and dried out with some spectrophotometric tissues. PLS models were built
16 on both the whole NIR spectral region (400 and 2500 nm) and on selected interval related to
17 phosphates (1600-1900 nm) and ammonium (2050-2300 nm).

¹ The electromagnetic spectrum is traditionally sectioned with reference to wavelength (nm) or wavenumber (cm^{-1}). The spectral region related to Mid-Infrared is 2500 – 25000 nm (4000 – 400 cm^{-1}) and the one related to Near-Infrared is 800 – 2500 nm (12500 – 4000 cm^{-1}). The IR instrument used was an FT-based instrument, and, thus, inherently produced data linear in frequency (constant frequency spacing between data points): these spectra were therefore measured in wavenumbers, through this study. The NIR instrument used in this study was a diffraction grating-based instrument and, therefore, inherently produced data linear in wavelength (constant wavelength spacing between data points): Hence, the spectral units used for the NIR spectra in this work were nanometers. The reader should be aware of the different units used for the two different spectral ranges.

1 DATA ANALYSIS

2 All data were imported by MatLab 8.3.0.532 (MathWorks, Natick, MA) and any further elaboration
3 was performed there using both in-house scripts and functions and the PLS toolbox 7.8
4 (Eigenvector Research, Inc., Manson, WA).

5 Among the whole set of 25 samples for each of the matrices, 8 samples were selected to be used as
6 a validation set. They are written in red color in Figure 1 and Table S1 and additionally marked
7 with a circle in Figure 1. These samples were therefore left out from the phase of building the
8 calibration models and they were subsequently used to check the performances of this model on
9 “unknown” samples.

10 The Partial Least Squares models (PLS) [33-35] have been assessed and compared in their
11 predictive performances mainly based on their Prediction Error in %. The prediction errors have
12 been computed according to the statistical calculations reported in Skou et al. [36]. The prediction
13 uncertainty ($s_{PE_i}^2$) of each validation sample i , was first calculated as:

$$14 \quad s_{PE_i}^2 = \left(\frac{1}{N} + h_i \right) \cdot \text{MSEC} + \text{MSEC}$$

15 where N represents the number of samples in the calibration set; h_i is the leverage of the i th
16 validation sample; and MSEC (mean squared error in calibration) is the observed error of the PLS
17 calibration model. Then the variance obtained was converted in standard deviation (s_{PE_i}) by taking
18 the square root and averaged to obtain the average prediction error over the validation samples.
19 Finally, this quantity was expressed in % by dividing it by the range of concentrations of the models
20 (90 mM) and multiplying it by 100.

21 Other parameters used through this paper to evaluate the PLS models and mentioned in the Figures
22 are the typical indicators of a PLS model, namely: number of latent variables (is the number of PLS

1 components used to build the model), RMSE (Root Mean Squared Error, corresponds to the random
2 uncertainty of the model), Bias (corresponds to the systematic error of the model), R^2 (coefficient of
3 determination). All these parameters are calculated and reported in Calibration (thus, the set of
4 samples used to build the models), Cross-Validation (CV), and in Prediction (the set of samples
5 used to validate the models).

6 **RESULTS AND DISCUSSION**

7 The three sets of samples prepared in pure water, minimal medium and complex medium,
8 respectively, were analyzed by two different vibrational spectroscopic techniques (MIR and NIR)
9 and by the two colorimetric assays for phosphates and ammonium content, used as reference
10 methods. The absolute quantification of phosphates was also obtained by means of ^{31}P NMR
11 spectroscopy. This approach is highly selective as only species containing phosphorous are
12 detected. Furthermore, only a simple sample preparation (addition of D_2O) is required, which
13 ensure a higher reliability of this method as a reference analysis. An accurate analysis and
14 comparison of all the reference methods was performed and it is reported in Appendix A

15 Reference measurements comparison, together with a detailed inspection of the error source for the
16 phosphate assay. Given the higher uncertainty (see Fig. A1, standard deviations values) in the
17 reference values of phosphates obtained by the colorimetric kit, with respect to the one obtained by
18 NMR ($< 2.4 \text{ mM}$), and the better agreement of the NMR results with the theoretical values, the
19 following regression models for phosphates were built on the NMR results. The ammonium models
20 were built on the colorimetric kit values that were considered satisfactory. We want to strike here
21 that the choice of reference method is as important as the calibration process itself, as the model can
22 never be better as the reference method.

1 INFRARED SPECTROSCOPY

2 The mid IR spectra were investigated at first for the potential of simultaneous analysis of
3 phosphates and ammonium as this type of spectra is in general more informative than NIR.

4 **Raw spectra inspection and interpretation.** The raw IR spectra of the three sets of solutions,
5 shown in Figure 2, were inspected to verify if quantification of the salts under study was possible
6 and to identify the regions where these salts had signals. Once the spectra were colored according to
7 the concentration of phosphates and ammonium, two specific spectral regions were identified as
8 concentration dependent (Figure 2B, C, E, F, H, I). The intense peak centered at 1636 cm^{-1}
9 dominating all spectra is due to the water that is the main component in all the solutions. The P-O
10 stretching region ($800\text{-}1400\text{ cm}^{-1}$ [37]) displayed several peaks, which systematically changed
11 intensity according to the amount of phosphates in the solution for all three different sets of
12 synthetic samples (Figure 2, inserts B, E and H). These are due to the symmetric and antisymmetric
13 stretching modes of the P-OH and P-O₂ bonds in HPO_4^{2-} and H_2PO_4^- , respectively, and their
14 complexes [37-39]. The shape and the position of these peaks varied among the three, different
15 media due to the differences in the compositions of the solutions. Despite these differences, the
16 dependence of the absorbance in this region from the amount of phosphates added in the samples
17 was evident. Figure 2C shows a peak with maximum intensity around 1455 cm^{-1} ; this can be
18 assigned to the triply degenerate bending mode of the tetrahedral NH_4^+ ion [40, 41]. The
19 dependence of the absorbance of this peak from the amount of salt added was evident for all three
20 sets of samples, despite the low intensity.

21

22 **Figure 2. Raw Infrared spectra of (A) water, (D) minimal media and (G) complex media**
23 **solutions.**

1

2 **PLS Modeling.** To determine the concentration of the two compounds in the samples, the data were
3 then analyzed by Partial Least Squares regression (PLS) [33-35]. For all the three sets of synthetic
4 samples two different PLS1 models were built for phosphates and ammonium concentrations,
5 respectively. The spectra were pre-processed by MSC (Multiplicative scatter Correction [42, 43])
6 and mean-centering, and different models were built on the whole spectral region (800 – 2250 cm^{-1})
7 as well as on the region related to the target species (950 – 1200 cm^{-1} for phosphates and 1360 –
8 1540 cm^{-1} for ammonium).

9 For clarity, the results are described with focus on the data set describing the complex medium
10 samples (Figure 3), as they resemble the fermentation matrix best. The results of the other data sets,
11 describing the water and minimal medium samples, can be regarded as similar and can be found
12 summarized in Supplementary Material, Table S4, as well as in Figure 4 at the end of this section.

13 The results obtained for the entire spectral region (800 – 2250 cm^{-1}) are shown in Figure 3. Both
14 models were successful using 3 and 4 component and resulting in a RMSEP of 5.6 and 2.5 mM,
15 respectively (Figure 3), corresponding to a prediction error of 3.63% and 3.15% (Figure 4A and B).

16 Inspection of the Variable Importance in Prediction (VIP) index [44] and regression coefficients for
17 the models of each of the species strongly highlighted one specific peak as the most important to
18 build the model (selected bands in the Figure). The two selected bands are coherent with the
19 respective regions identified and known in literature to be correlated with specific vibrational
20 modes of the phosphates and ammonium ions, namely the peak at 1076 cm^{-1} for the phosphates and
21 the one at 1454 cm^{-1} for the ammonium. This finding confirm that the model is built upon direct
22 correlations with relevant wavelength and it can be considered trustworthy and robust.

23

1 **Figure 3. PLS models on complex media, whole IR region.**

2

3 **Figure 4. Comparison of all PLS models on IR (A and B) and NIR (C and D) data by their**
4 **relative Prediction Error in %.**

5

6 Focusing only on the interval related to phosphates and ammonium, highlighted in Figure 3, new
7 PLS models were computed on only these selected intervals. A comparison between the models
8 built on the full region ($800 - 2250\text{ cm}^{-1}$) and on the region related to the target species ($950 - 1200$
9 cm^{-1} for phosphates (Figure 3 panels A, B and C) and $1360 - 1540\text{ cm}^{-1}$ for ammonium (Figure 3
10 panels D, E and F)) regarding their most important model parameters is presented in Figure 4A and
11 B (and in Supplementary Material, Table S4). In Figure 4 the results of the prediction errors for all
12 models (full and selected region for phosphates and ammonium) are summarized and compared.

13 The first and most general consideration that could be drawn from Figure 4A and B was that overall
14 the errors in the models increased when passing from water to minimal and complex medium due to
15 increased complexity of the matrix. This behavior was expected due to the higher number of
16 interfering species present in the complex matrix, that can hamper the model or that might react
17 with the analytes in different ways altering their spectral signals or their response to the kits.

18 The VIP indices in Figure 3, showed only a very narrow region of the spectra having very high
19 values, whereas the rest of the spectra had VIP indices below 1. This led to the hypothesis that
20 models constructed on only these regions, would not change, or would slightly increase the
21 performances of the PLS models. This showed to be only partially correct, according to Figure 4
22 and Supplementary Material (Table S4) as many of the models computed on the selected region

1 showed quite similar, but slightly poorer performances than the corresponding models on the whole
2 region. This behavior suggests that, even when the VIP index of the major part of the spectra is
3 below, or very close to 1, the presence of these areas in the model contribute to its stabilization,
4 especially helping to improve its selectivity towards the target species. To further investigate this,
5 the prediction error was investigated in more detail. The result of the analysis (described in
6 Appendix B) confirmed the proposed hypothesis showing that the intervals (for both analytes) were
7 less sensitive than the whole region.

8 NEAR INFRARED SPECTROSCOPY

9 Although providing less detailed information than the IR spectra, a NIR on-line monitoring probe
10 would be the easiest measurement to implement for a fermentation process. This technique was
11 therefore investigated as well.

12 **Raw spectra inspection and interpretation.** As done before for the IR spectra, the raw NIR
13 spectra of the three sets of solutions were visually inspected by overlaying of spectra to verify if a
14 quantification of phosphates and ammonium was possible. The identification of any spectral region
15 that varied according to the content of these analytes was not easy to see by naked eye, so the
16 correlation coefficient of each spectral wavelength with the reference values of each analyte was
17 computed (Supplementary Material, Figure S1). The spectral regions with the highest correlation to
18 the concentration of the analytes were the same in the three batches of samples. These spectral
19 regions correspond to bands reported in the literature for the two compounds under study. The
20 regions correlated with phosphates could be attributed to the third overtones of the P–O bonds in
21 phosphates salts, that are reported between 1600 and 2000 nm [45]. In general, phosphates and
22 other inorganic P compounds are hardly detectable by NIR spectroscopy due to the weak P–O
23 dipole moment [46, 47]. Consequently, the correlation was less prominent for phosphates, with a

1 correlation factor < 0.7 (see Figure S1 in Supplementary Material), while it was around 0.8 for
2 ammonium. Bands between 2100 and 2400 nm (attributed to the second overtones of the N-H
3 bonds in NH_4^+ [45]) have been used more than once for determination of ammonium in samples
4 similar to the ones under study, for examples in references [16, 48, 49], and other examples are
5 listed in [11]. However, the main purpose of the correlation analysis was to verify if any spectral
6 region could be identified easily as respondent of phosphates or ammonium, confirming the
7 attribution of the peaks hypothesized based on the literature.

8 **PLS modeling.** All the spectra were pre-processed by mean-centering, that was chosen over
9 Standard Normal Variate (SNV), and first and second derivative because led to better results (data
10 not shown). Then, for each set of samples, two different PLS models, for phosphates and
11 ammonium respectively, were built on the whole NIR spectral region, as was done with IR spectra.

12 Inspection of the results showed that the PLS models based on the NIR data of minimal and
13 complex media for both analytes allowed a good quantification of the compounds with errors in
14 prediction between 3.1 and 7.1% (Supplementary Material, Table S5). As an example, Figure 5
15 presents the models for phosphates and ammonium in complex medium while a summary of the
16 performances of all the models is presented in Figure 4. Investigation of the VIP index and of the
17 regression coefficient of the models indicated the spectral regions 1600 – 2000 nm for phosphates
18 and 2100-2400 for ammonium as particularly important for building the respective models (Figure
19 5B, C, E, F). This improved our understanding of the predictions: while for ammonium prediction,
20 a true correlation between the most important variables (wavelengths) and the analyte of interest is
21 found (VIP scores and regression coefficients are especially high and positive within the expected
22 important spectral region, highlighted in blue in Figure 7 E and F), the phosphates prediction
23 strongly suggests that the model is based on an inverse correlation with the amount of water, and it
24 is therefore not truly related to the amount of phosphates itself. This can be concluded from the

1 region around 1900 nm corresponding to water (Figure 5B and C): the high and positive VIP scores
2 in that region indicate these variables as particularly important for the model, while the
3 correspondent high and negative regression coefficients indicate a negative correlation between this
4 region and the analyte. Taking also into account that this region with high VIP scores mostly lies
5 outside the expected important spectral region for phosphates (highlighted in blue in the figure), this
6 suggests an inverse relation between the absorbance in that region and the phosphates content.

7 To force all the models to focus on the spectral regions with the highest correlation to the two
8 analytes, a new set of models was built using only the most relevant intervals for the two analytes.
9 These intervals were selected based on the literature and on the inspection of the correlation
10 coefficient pictured in Figure S1, and were, in fact, 1600-1900 nm for phosphates and 2050-2300
11 nm for ammonium. It has to be stressed here that, for phosphates, the selected regions intentionally
12 excluded the peak of water and therefore do not match with the areas of highest VIP scores showed
13 previously in Figure 3. The performance of these last models based on the prediction error [36],
14 together with the ones on the whole spectra, are shown in Figure 4 and summarized in
15 Supplementary Material, Table S5.

17 **Figure 5. PLS models on complex media, whole NIR region.**

19 The models on the intervals showed similar performances to the models on the entire spectral
20 region, with prediction errors always of similar magnitude, except for the models in water (Figure
21 4). As an example, the models on the intervals for the complex media samples are shown in
22 Supplementary Material, Figure S2. As highlighted in Figure 5, the model for ammonium focused
23 on the peak reported in the literature and highlighted in Figure S1 as well, whereas the one for

1 phosphates was based only on the tail of the peak at 1900 nm and with a negative correlation
2 coefficient.

3 The model for phosphates (and to a lower extent the one for ammonium) in water had an extremely
4 high error compared to the other models (Figure 4). Based on a more detailed investigation of the
5 prediction error contributions (reported in Appendix B), this was attributed to an error of the model
6 itself, resulting from deviation from the Lambert-Beer law, non-linearity, or in more general terms,
7 lack of information in the spectral data. This supported the previous hypothesis of lack of
8 information in NIR data (built of the model on indirect correlations) in particular with respect to
9 phosphate and to a lower degree for ammonium. Apparently, a more complex background than
10 water is needed to gather spectral data rich enough in information to build a PLS model. This is
11 confirmed by the certainly decreased error for NIR models built on the samples of minimal and
12 complex medium (see Appendix B for more details). Moreover, a comparison in sensitivity of the
13 techniques (Fig. B1 and Supplementary Material, Table S3), highlighted a much lower sensitivity of
14 NIR spectroscopy towards both analytes. Comparing all the best models (selected among the ones
15 built on the whole spectra or on the intervals) on IR and NIR data (Table S6, Supplementary
16 Material), we could observe that the best models were always the one built on the IR spectra, that
17 had overall the lowest RMSEP with a lower number of latent variables. This is not surprising
18 considering the lower sensitivity of the NIR mentioned previously and the hypothesis that most
19 models built on NIR spectra were built on indirect correlations. In this respect we have to mention
20 that the fact of basing a model on indirect correlations does not, per se, hamper the possibility of
21 using the model: we observed in Table S6, for example, that some of the NIR models are similar in
22 performances to the ones built on IR and could potentially be used to predict the analytes. In cases
23 of models based on indirect correlations, though, the user has to keep in mind that the persistence of
24 such correlation becomes an essential condition for the future applicability of the model. If, for any

reason, any change of the matrix or any change in the fermentation conditions occurs, this could potentially break the correlation making the model inaccurate. Hence, a frequent and accurate procedure of calibration maintenance becomes of capital importance.

IR-SPECTROSCOPY – PLS MODELS OF ALL SAMPLES

In order to finally prove the reliability of our approach some last models have been computed where the samples of the three different kinds of solutions have been considered all together. These models were computed for both analytes, both, on the whole spectra and on the selected region based on the IR data as shown to be superior compared to the NIR results. The results of the models are summarized in Supplementary Material, Table S7 and in Figure 4E.

With respect to Figure 4E the overall model for ammonium showed a slightly higher error than the model for phosphates and this could be attributed to the lower sensitivity of the technique for the ammonia peak (resulting in very low intensity of the peak), compared to the peak for phosphates (Figure 2B and C) (for further details and explanation on this, see Appendix B, Fig. B2 and related text). Figure 6 shows the model built on all IR spectra acquired in the three different media built on the whole region. Focusing on the models built on the complex medium samples representing the approach closest to a real fermentation medium, a comparison of the models in Figure 4A and B with Figure 4E shows how similar both approaches perform in general. The similar performance independent of the data set used (either combining the three data sets or keeping them separately) is further represented by comparing the error of prediction in Figure 6 and Figure 3 which is in both cases around 3 and 4 %. This confirms that our strategy is solid even in cases where different batches of samples are analyzed together, given that all kind of samples are present in the calibration set.

1

2 **Figure 6. PLS models on three sets of samples, whole IR region (800 – 2250 cm⁻¹).**

3

4 Inspecting Figure 6B, C in more detail it can be seen that a broader spectral region was required to
5 build the PLS models for phosphates than the corresponding one in Figure 3, extending the spectral
6 range found by visual inspection and according to literature (highlighted in blue). Additionally, with
7 respect to the VIP index in Figure 3), Figure 6B shows a more complex pattern in the respective
8 spectral region. This confirmed the complex nature of the spectral region associated with
9 phosphates and its strong dependence on the medium. Contrarily, the model for ammonium
10 presented a pattern of VIP and regression coefficients (Figure 6E and F) that resemble completely
11 the respective ones in Figure 3E and F suggesting a very strong correlation in the particular region
12 with the analyte.

13 **CONCLUSION**

14 In this paper different spectroscopic techniques were evaluated with respect to their ability to
15 quickly determine the content of phosphates and ammonium in a wide range of concentration in
16 fermentation media with a minimum of sample preparation.

17 Due to their potential for on-line applications the performance of NIR and IR spectroscopy were
18 explored in combination with PLS modeling. As reference methods a traditional colorimetric assay
19 kit was chosen for ammonium, while ³¹P NMR measurement was chosen for phosphates, as it was
20 found more reliable than the colorimetric kit for phosphates tested in comparison.

21 The PLS models based on MIR spectroscopy gave good results, with prediction errors lower than
22 5% for both analytes in the datasets of the complex and minimal media samples, and even better

1 results for the one in water (with an error of about 2%). When PLS models were built on the spectra
2 of all types of samples, the errors remained stable between 3.5 and 5% for both analytes, confirming
3 the generalizability of this approach.

4 The PLS models based on NIR spectroscopy, on the other hand, provided acceptable results for the
5 concentration of ammonium in some of the media, and satisfactory but indirect models for
6 quantification of phosphates in the same medium. This method is therefore considered inferior to
7 the previous one for the task under exam, even if, given the much easier applicability of the
8 technique on-line, it could be still used under very controlled circumstances and following an
9 accurate routine of calibration maintenance.

10 Overall, the results presented in this paper are very promising in terms of fast and accurate analysis
11 of phosphates and ammonium in fermentation samples, especially considering that the models are
12 built on experimental designed samples where the two compounds were uncorrelated and particular
13 care was put into keeping the models free from dependence on any indirect or spurious correlation.

14

15 **GLOSSARY OF CHEMOMETRIC AND SPECTROSCOPIC TERMS**

16 Chemometrics: scientific discipline of extracting information from chemical data by means of
17 mathematical and statistical models.

18 Multivariate data analysis: Statistical and mathematical approach to data analysis that implies the
19 simultaneous use of several measured variables on the same set of samples.

20 Partial Least Square (PLS): regression models: one of the most general and common regression
21 models in chemometrics. For a detailed explanation of its mathematical description see the
22 reference cited in the text [33-35].

1 Cross Validation (CV): iterative process used as internal validation and to select the most
2 appropriate parameters for the chemometric model. Iteratively, predefined data set segments are left
3 out from the modelling phase and are used as “external samples” to evaluate the performance on the
4 model. These predictions are used against the reference values to estimate the RMSECV.

5 Root mean Square Error (RMSE): corresponds to the random uncertainty of the model.

6 Coefficient of determination (R^2): proportion of the variance in the dependent variable (y) that is
7 predictable from the independent variables (X).

8 Variable Importance in Prediction (VIP): index associated to each variable and explaining its
9 importance in the chemometric model. For a detailed explanation of its mathematical description
10 see the reference cited in the text [44].

11 Bias: corresponds to the systematic error in the model.

12 Regression Coefficients: vector of proportionality constants used to describe the relations between
13 the dependent variable (y) and the independent variables (X). For a detailed explanation of its
14 mathematical description see the reference cited in the text [33-35].

15 Leverage: diagnostic measure describing the influence of each sample (or each variable) used for
16 calibration, on the mathematical model.

17 Full factorial design: (or fully crossed design) type of experimental design that includes
18 experiments with all possible combinations of all the factors considered in the design at all their
19 possible levels.

20 Attenuated total reflectance (ATR): sampling spectroscopic technique that enables the analysis of
21 solid or liquid samples without further preparation or isolation of their components. For a detailed
22 explanation see specialized literature [50].

Mean centering (preprocessing): Variable pre-processing consisting of subtracting by each variable their mean over all the samples. For a detailed explanation of its mathematical description see the reference cited in the text [51].

Standard Normal variate (SNV): Variable pre-processing for scatter correction consisting of for each sample, subtracting each variable of the mean of all the variables for that sample and dividing by the standard deviations of all the variables for the same sample. For a detailed explanation of its mathematical description see the reference cited in the text [51].

Multiplicative Scatter Correction (MSC): Variable pre-processing for scatter correction consisting of rescaling and shifting of the spectra according to scalar parameters found by plotting each spectrum towards a reference spectrum. For a detailed explanation of its mathematical description see the reference cited in the text [51].

APPENDIX A

REFERENCE MEASUREMENTS COMPARISON

The results obtained with the reference measurements on each sample for phosphates and ammonium are reported in Fig. A1 and Table S1, together with the theoretical values calculated for the preparation of the respective sample. Clearly, the ammonium assay performed better than the phosphates assay with respect to standard deviation (indicated as STD in absolute values, in Fig. A1) within the same sample as well as with respect to the desired set point. Overall (considering also differences in the three different media) the standard deviation was found to be < 1.6 within the ammonium assay and > 4 for the phosphate assay. Compared to the ammonium assay, in practice the phosphates assay is highly dependent on a calibration curve that must be prepared each time the assay is run to determine the background value. Consequently, for the phosphates determination in water, minimal and complex medium three different calibration curves were

1 acquired causing the differences between the three data sets. Moreover, for the phosphate assay the
2 concentration range recommended (in the order of nM) was much lower than the one needed for the
3 ammonium assay (mM), and the samples had to be diluted up to 6000 times, while they were diluted only up
4 to 20 times for the ammonium assay. As the final dilution was achieved through a dilution series, the
5 variation in the higher concentrated phosphates samples can be referred to the higher number of dilution
6 steps necessary to achieve the desired concentration, each time introducing a small operational error. This
7 was concluded after comparing the different data sets. Subsequently it was shown that even for higher
8 concentrated samples in complex medium the average standard deviation can be reduced (more than bisected
9 compared to the standard deviation stated in Fig. A1 for the phosphate determination in complex medium by
10 means of the kit) by choosing less dilution steps (maximum 2 instead of up to 4). As the standard deviation is
11 not only increasing for higher concentrated samples but also from water to minimal to complex medium (Fig.
12 A1 and Table S1), one could also argue for interferences within the assay arising with increased medium
13 complexity. Although the supplier does not state the method principle for the phosphate assay used, the
14 malachite green method for phosphate determination can be assumed due to the formation of a green color,
15 the wavelength used for quantification (650 nm) and the detection range in nanomolar scale. The
16 spectrophotometric malachite green method for determining inorganic phosphate is well established, but the
17 final performance depends on a variety of reaction conditions broadly optimized in literature [52-59]. As the
18 exact assay composition is unknown, a discussion about possible interferences can only take place on a
19 general level. Possible interferences can be introduced by the presence of organic phosphate species, salts or
20 organic acids. Organic phosphate species can generally be prone to hydrolysis under acidic conditions
21 assumed to be found in the assay resulting in additional free inorganic phosphate. A salt error can be
22 introduced by CaCl_2 , but it was shown by [60] that no interference with the assay was observed up to 40 mM
23 CaCl_2 . The salts of organic acids able to coordinate to the complex (responsible for the green color
24 formation) can increase the signal, but also here concentrations first interfere over 200 mM [52] which does
25 not hold for our samples. Finally, no interferences are mentioned by the supplier. Based on our results, the
26 presence of the interfering species mentioned are expected to be of minor contribution compared to the
27 operational error introduced by several dilution steps. This was observed by a higher standard deviation for

the samples in the upper end of the concentration investigated. These samples required the highest dilution and contained only traces of the interfering species.

Fig. A1: Comparison of the reference analysis and theoretical values. Comparison of the theoretical values of phosphates with the colorimetric analysis values (a) and with the ^{31}P NMR values (b) in the three, different media under study: water, minimal and complex medium. Comparison of the theoretical values of ammonium with the colorimetric analysis values (c).

APPENDIX B

BREAK DOWN OF THE PREDICTION ERRORS CONTRIBUTIONS

To identify the origin of the prediction error and better compare the different PLS models, the prediction error was broken down into its contributions, following the calculation reported in the literature [36, 61].

The main contributors to the final error in a PLS model can be distinguished as *error on the y* (namely, errors on the reference values used to calibrate and validate the models), *error on the X* (namely, errors on the spectral data acquired on the samples), and *error on the model* itself (part of the error unrelated to X and y that can be due to, e.g., deviation from Beer-Lambert law, interferences, operational errors or shifts in the spectra or other deviations). The contributions to the prediction error due to the errors in the X and y (the spectra and the reference values, respectively), and to the model error itself are shown Fig. B1 and Tabel S2. The *error in X* was calculated multiplying the square of the norm of the regression vector of the model **b** by the variation in X [36]. The *error in y* corresponds to the variance of the reference measurements, and it was calculated by the standard deviation of the replicates performed. The *model error* was then obtained by subtracting the two above mentioned errors (in X and y) from the observed error of the calibration model (MSEC), that is computed as the square of the RMSEC.

By inspection of Fig. B1, a first general observation is that the error on the y was generally higher than the overall error on the X, and therefore a big part of the prediction error resulted from the uncertainty in the

1 reference method. The figure shows, further, that the uncertainty related to the measurement of phosphates
2 was always higher than the one for ammonium and this difference was particularly high for the minimal
3 medium samples, where the variance of phosphates measurement was even higher than in the case of the
4 complex medium samples.

5 A comparison between the results of the NIR and IR models (Figure 5 in the main text) reveals that the
6 errors for the NIR models were in general higher than the ones for IR. The biggest differences, though, were
7 the prediction errors of the NIR models in water. They reached up to 27 % for phosphate and up to 17 % for
8 ammonium while all the others models, IR and NIR for both analytes, only showed an error of maximal 7 %.
9 Since the reference values were the same for both NIR and IR, the cause of the extremely high errors in
10 water for the NIR spectral data had to be linked either to the **X** block or the to the models themselves.

11 Inspection of Fig. B1 clearly demonstrates that the high error observed for the NIR models of phosphate and
12 ammonium built on the water samples (Figure 5, main text) resulted from the model error itself (Fig. B1 and
13 Table S2). The respective model error was huge, around 200 and 500 mM² for phosphate and ammonium,
14 respectively, while the error related to **X** for both spectroscopies could be neglected. At the same time, Fig.
15 B1 and Table S2 show that the NIR model errors in water were over 30 times higher than the average error
16 of all the other IR and NIR models for both analytes (around 15 mM²). They could have been due to
17 deviation from the Lambert-Beer law, non-linearity, or in more general terms, lack of information in the
18 spectral data. Considering all the above and keeping in mind that we observed high positive VIP scores but
19 high negative regression coefficients for the phosphates model built on NIR data (see Results and
20 discussion) we could argue that NIR models for phosphate were not robust, as they were built upon indirect
21 correlations. Hence, the NIR technique has to be considered inferior to the IR one for analyzing both analytes
22 at the same time.

Fig. B1: Contribution to the Prediction Error given by the Error in X (spectra) and y (reference measurements), for both IR and NIR spectra and for both analytes.

In order to further improve the understanding of the different performance of the two spectroscopic techniques, the error on the X was broken down into its components: namely, the *spectral variance* and the *(lack of) sensitivity in the spectral signal*. The spectral variance refers to the deviation between different replicates of the same sample and, in practice, describes the variability of the spectral measurement with respect to the analyte.

The sensitivity of the spectral signal can be quantified in the, so-called, *net analyte signal* (NET). The NET corresponds to the part of the spectrum related to the pure analyte, thus uncorrelated to the response of all other compounds. Mathematically the NET is inversely proportional to the size of the Euclidean norm of the regression coefficient vector (\mathbf{b}) of the PLS models [62, 63]. The norm of \mathbf{b} represents therefore the *(lack of) sensitivity* of the measurement with respect to the specific analyte [62]. In practice this means that the higher the value of norm of \mathbf{b} , the less sensitive the measurements are with respect to the analyte under study. The results are shown in Table S3 and Fig. B2.

Fig. B2: Interpretation of the Error in X (spectra) by its contributions: Variance in X and norm of \mathbf{b} .

Regarding the variance in X , the most evident observation is that the variability of the whole spectral region was generally higher than (or equal to) the variability of the spectral intervals for the two analytes. The only exception to this behavior is shown by the IR spectra for prediction of phosphates content in the complex medium. This deviation might be explained by the increased complexity of the matrix causing additional variations in the spectral region related to phosphates caused by other organic components like sugars. Overall, focusing only on the variability of the models on the intervals, it was demonstrated that the highest

variability was present in the region related to ammonium for the NIR region, and to the one of phosphates for the IR region.

The norm of **b** showed consistently a higher value for all the models on the intervals compared with the models built on the whole spectrum. As already explained, the higher the value of norm of **b**, the less sensitive is the measures with respect to the analyte under study. This showed that the intervals (for both analytes) were less sensitive than the whole region and confirmed our original hypothesis that using the whole spectra contributes to stabilize the models, overcoming interferences problems. This effect is present in both IR and NIR spectra but it is much more obvious for the NIR models. Examining the same plots, it can also be noticed that the values of the norm of **b** are generally much higher for the NIR technique than for the IR, and particularly high for the norm of **b** on the intervals, indicating a lower sensitivity of NIR spectroscopy towards both analytes. Finally, when comparing all the models for the two analytes, a similar level of sensitivity was observed for both analytes in the NIR models, while a higher sensitivity was found for the phosphates models on the IR models. This was explained by the higher intensity of the phosphates peaks compared to the ammonium peak in the IR spectrum, which makes the spectrum more sensitive towards this analyte. Notwithstanding this, the overall errors in the models for phosphates were higher than the ones for ammonium (Figure 5, in the main text) and this difference could be attributed to the higher error on the reference measurement for phosphates (Fig. B1).

FUNDING:

This work was supported by the BIOPRO project (www.biopro.nu) thanks to the funds granted by Innovationsfondens [Grant Number 4105-00020B]. The funding source had no involvement in the study design, collection, analysis, interpretation of data, writing of the manuscript, nor in the decision to submit the article for publication.

1 REFERENCES

- [1] Dawson CJ, Hilton J, Fertiliser availability in a resource-limited world: Production and recycling of nitrogen and phosphorus. *Food Policy*, 2011; 36: S14-S22. <http://dx.doi.org/10.1016/j.foodpol.2010.11.012>.
- [2] Schoumans OF, Bouraoui F, Kabbe C, Oenema O, van Dijk KC, Phosphorus management in Europe in a changing world. *AMBIO*, 2015; 44: 180-192. 10.1007/s13280-014-0613-9.
- [3] Garn M, Gisin M, Thommen C, Cevey P, A flow injection analysis system for fermentation monitoring and control. *Biotechnol Bioeng*, 1989; 34: 423-8. 10.1002/bit.260340402.
- [4] Landgrebe D, Haake C, Höpfner T, Beutel S, Hitzmann B, Scheper T, et al., On-line infrared spectroscopy for bioprocess monitoring. *Applied Microbiology and Biotechnology*, 2010; 88: 11-22. 10.1007/s00253-010-2743-8.
- [5] Vojinović V, Cabral JMS, Fonseca LP, Real-time bioprocess monitoring: Part I: In situ sensors. *Sensors and Actuators B: Chemical*, 2006; 114: 1083-1091. <https://doi.org/10.1016/j.snb.2005.07.059>.
- [6] Schmidt WJ, Meyer HD, Schügerl K, Kuhlmann W, Bellgardt KH, On-line analysis of fermentation media. *Anal Chim Acta*, 1984; 163: 101-109. [http://dx.doi.org/10.1016/S0003-2670\(00\)81498-1](http://dx.doi.org/10.1016/S0003-2670(00)81498-1).
- [7] Lukkari I, Ruzicka J, Christian GD, Determination of total ammonium-nitrogen and free ammonia in a fermentation medium by sequential injection analysis. *Fresenius' Journal of Analytical Chemistry*, 1993; 346: 813-818. 10.1007/bf00321296.
- [8] Lynggaard-Jensen A, Eisum NH, Rasmussen I, Svankj Jacobsen H, Stenstrøm T, Description and test of a new generation of nutrient sensors. *Water Science and Technology*, 1996; 33: 25-35. [http://dx.doi.org/10.1016/0273-1223\(96\)00156-4](http://dx.doi.org/10.1016/0273-1223(96)00156-4).
- [9] Teixeira AP, Oliveira R, Alves PM, Carrondo MJT, Advances in on-line monitoring and control of mammalian cell cultures: Supporting the PAT initiative. *Biotechnology Advances*, 2009; 27: 726-732. <http://dx.doi.org/10.1016/j.biotechadv.2009.05.003>.
- [10] Wang RY, Jarratt JA, Keay PJ, Hawkes JJ, Coakley WT, Development of an automated on-line analysis system using flow injection, ultrasound filtration and CCD detection. *Talanta*, 2000; 52: 129-139. [http://dx.doi.org/10.1016/S0039-9140\(99\)00342-2](http://dx.doi.org/10.1016/S0039-9140(99)00342-2).
- [11] Cervera AE, Petersen N, Lantz AE, Larsen A, Gernaey KV, Application of near-infrared spectroscopy for monitoring and control of cell culture and fermentation. *Biotechnology progress*, 2009; 25: 1561-1581.
- [12] Kornmann H, Valentinotti S, Duboc P, Marison I, von Stockar U, Monitoring and control of *Gluconacetobacter xylinus* fed-batch cultures using in situ mid-IR spectroscopy. *Journal of Biotechnology*, 2004; 113: 231-245. <http://dx.doi.org/10.1016/j.jbiotec.2004.03.029>.
- [13] Pollard D, Buccino R, Connors N, Kirschner T, Olewinski R, Saini K, et al., Real-time analyte monitoring of a fungal fermentation, at pilot scale, using in situ mid-infrared spectroscopy. *Bioprocess and Biosystems Engineering*, 2001; 24: 13-24. 10.1007/s004490100226.
- [14] Triadaphillou S, Martin E, Montague G, Norden A, Jeffkins P, Stimpson S, Fermentation process tracking through enhanced spectral calibration modeling. *Biotechnology and Bioengineering*, 2007; 97: 554-567. 10.1002/bit.21248.
- [15] Tamburini E, Marchetti MG, Pedrini P, Monitoring Key Parameters in Bioprocesses Using Near-Infrared Technology. *Sensors (Basel, Switzerland)*, 2014; 14: 18941-18959. 10.3390/s141018941.
- [16] McShane MJ, Cote GL, Near-Infrared Spectroscopy for Determination of Glucose, Lactate, and Ammonia in Cell Culture Media. *Appl. Spectrosc.*, 1998; 52: 1073-1078.
- [17] Lewis CB, McNichols RJ, Gowda A, Cote GL, Investigation of Near-Infrared Spectroscopy for Periodic Determination of Glucose in Cell Culture Media in Situ. *Appl. Spectrosc.*, 2000; 54: 1453-1457.
- [18] Riley MR, Arnold MA, Murhammer DW, Matrix-Enhanced Calibration Procedure for Multivariate Calibration Models with Near-Infrared Spectra. *Appl. Spectrosc.*, 1998; 52: 1339-1347.

- 1 [19] ASTM I. ASTM E1655 - 05(2012) Standard Practices for Infrared Multivariate Quantitative Analysis.
2 2012, ASTM International.
- 3 [20] Hakemeyer C, Strauss U, Werz S, Jose GE, Folque F, Menezes JC, At-line NIR spectroscopy as
4 effective PAT monitoring technique in Mab cultivations during process development and
5 manufacturing. *Talanta*, 2012; 90: 12-21. <https://doi.org/10.1016/j.talanta.2011.12.042>.
- 6 [21] Arnold SA, Crowley J, Woods N, Harvey LM, McNeil B, In-situ near infrared spectroscopy to monitor
7 key analytes in mammalian cell cultivation. *Biotechnol Bioeng*, 2003; 84: 13-9. 10.1002/bit.10738.
- 8 [22] Tamburini E, Vaccari G, Tosi S, Trilli A, Near-infrared spectroscopy: a tool for monitoring submerged
9 fermentation processes using an immersion optical-fiber probe. *Appl Spectrosc*, 2003; 57: 132-8.
10 10.1366/000370203321535024.
- 11 [23] Mazarevica G, Diewok J, Baena JR, Rosenberg E, Lendl B, On-Line Fermentation Monitoring by Mid-
12 Infrared Spectroscopy. *Appl. Spectrosc.*, 2004; 58: 804-810. 10.1366/0003702041389229.
- 13 [24] Rhiel M, Cohen MB, Murhammer DW, Arnold MA, Nondestructive near-infrared spectroscopic
14 measurement of multiple analytes in undiluted samples of serum-based cell culture media.
15 *Biotechnol Bioeng*, 2002; 77: 73-82.
- 16 [25] Finn B, Harvey LM, McNeil B, Near-infrared spectroscopic monitoring of biomass, glucose, ethanol
17 and protein content in a high cell density baker's yeast fed-batch bioprocess. *Yeast*, 2006; 23: 507-
18 517. 10.1002/yea.1371.
- 19 [26] Petersen N, Ödman P, Padrell AEC, Stocks S, Lantz AE, Gernaey KV, In situ near infrared
20 spectroscopy for analyte-specific monitoring of glucose and ammonium in streptomyces coelicolor
21 fermentations. *Biotechnology Progress*, 2010; 26: 263-271. 10.1002/btpr.288.
- 22 [27] Riley MR, Rhiel M, Zhou X, Arnold MA, Murhammer DW, Simultaneous measurement of glucose and
23 glutamine in insect cell culture media by near infrared spectroscopy. *Biotechnol Bioeng*, 1997; 55:
24 11-5. 10.1002/(sici)1097-0290(19970705)55:1<11::aid-bit2>3.0.co;2-#.
- 25 [28] Riley MR, Arnold MA, Murhammer DW, Wallis EL, Delacruz N, Adaptive calibration scheme for
26 quantification of nutrients and byproducts in insect cell bioreactors by near-infrared spectroscopy.
27 *Biotechnol Prog*, 1998; 14: 527-33. 10.1021/bp980022d.
- 28 [29] Gustafsson JP, Visual MINTEQ 3.0 user guide. KTH, Department of Land and Water Resources,
29 Stockholm, Sweden, 2011.
- 30 [30] Kazadi Mbamba C, Batstone DJ, Flores-Alsina X, Tait S, A generalised chemical precipitation
31 modelling approach in wastewater treatment applied to calcite. *Water Research*, 2015; 68: 342-
32 353. <https://doi.org/10.1016/j.watres.2014.10.011>.
- 33 [31] Kazadi Mbamba C, Tait S, Flores-Alsina X, Batstone DJ, A systematic study of multiple minerals
34 precipitation modelling in wastewater treatment. *Water Research*, 2015; 85: 359-370.
35 <https://doi.org/10.1016/j.watres.2015.08.041>.
- 36 [32] Christiansen T, Michaelsen S, Wümpelmann M, Nielsen J, Production of savinase and population
37 viability of *Bacillus clausii* during high-cell-density fed-batch cultivations. *Biotechnology and*
38 *Bioengineering*, 2003; 83: 344-352. 10.1002/bit.10675.
- 39 [33] Geladi P, Kowalski BR, Partial least-squares regression: a tutorial. *Anal Chim Acta*, 1986; 185: 1-17.
40 [https://doi.org/10.1016/0003-2670\(86\)80028-9](https://doi.org/10.1016/0003-2670(86)80028-9).
- 41 [34] Martens H, Næs T. Multivariate Calibration. In: B.R. Kowalski, Editor. *Chemometrics: Mathematics*
42 *and Statistics in Chemistry*, Dordrecht: Springer Netherlands; 1984, p. 147-156.
- 43 [35] Wold S, Martens H, Wold H. The multivariate calibration problem in chemistry solved by the PLS
44 method. In: B. Kågström and A. Ruhe, Editors. *Matrix Pencils: Proceedings of a Conference Held at*
45 *Pite Havsbad, Sweden, March 22-24, 1982*, Berlin, Heidelberg: Springer Berlin Heidelberg; 1983, p.
46 286-293.
- 47 [36] Skou PB, Berg TA, Aunsbjerg SD, Thaysen D, Rasmussen MA, van den Berg F, Monitoring Process
48 Water Quality Using Near Infrared Spectroscopy and Partial Least Squares Regression with
49 Prediction Uncertainty Estimation. *Appl Spectrosc*, 2017; 71: 410-421.
50 10.1177/0003702816654165.

- 1 [37] Rudolph WW, Raman- and infrared-spectroscopic investigations of dilute aqueous phosphoric acid
2 solutions. *Dalton Transactions*, 2010; 39: 9642-9653. [10.1039/C0DT00417K](https://doi.org/10.1039/C0DT00417K).
- 3 [38] Baril J, Max J-J, Chapados C, Titration infrarouge de l'acide phosphorique. *Canadian Journal of*
4 *Chemistry*, 2000; 78: 490-507. [10.1139/v00-038](https://doi.org/10.1139/v00-038).
- 5 [39] Klähn M, Mathias G, Kötting C, Nonella M, Schlitter J, Gerwert K, et al., IR Spectra of Phosphate Ions
6 in Aqueous Solution: Predictions of a DFT/MM Approach Compared with Observations. *The Journal*
7 *of Physical Chemistry A*, 2004; 108: 6186-6194. [10.1021/jp048617g](https://doi.org/10.1021/jp048617g).
- 8 [40] Max J-J, Chapados C, Aqueous ammonia and ammonium chloride hydrates: Principal infrared
9 spectra. *Journal of Molecular Structure*, 2013; 1046: 124-135.
10 <http://dx.doi.org/10.1016/j.molstruc.2013.04.045>.
- 11 [41] Wagner EL, Hornig DF, The Vibrational Spectra of Molecules and Complex Ions in Crystals IV.
12 Ammonium Bromide and Deutero-Ammonium Bromide. *The Journal of Chemical Physics*, 1950; 18:
13 305-312. [10.1063/1.1747623](https://doi.org/10.1063/1.1747623).
- 14 [42] Geladi P, MacDougall D, Martens H, Linearization and Scatter-Correction for Near-Infrared
15 Reflectance Spectra of Meat. *Appl. Spectrosc.*, 1985; 39: 491-500.
- 16 [43] Martens H, Jensen SA, Geladi P. Multivariate linearity transformations for near infrared reflectance
17 spectroscopy. in *Nordic Symposium on Applied Statistics*. 1983. Stokkland Forlag, Stavanger,
18 Norway.
- 19 [44] Wold S, Johansson E, Cocchi M, PLS-partial least squares projections to latent structures. 3D QSAR in
20 drug design, 1993; 1: 523-550.
- 21 [45] Williams P, Norris K. Near-infrared technology in the agricultural and food industries. St. Paul,
22 Minnesota: American Association of Cereal Chemists, Inc.; 1987.
- 23 [46] Malley DF, Yesmin L, Wray D, Edwards S, Application of near-infrared spectroscopy in analysis of soil
24 mineral nutrients. *Communications in Soil Science and Plant Analysis*, 1999; 30: 999-1012.
25 [10.1080/00103629909370263](https://doi.org/10.1080/00103629909370263).
- 26 [47] Niederberger J, Todt B, Boča A, Nitschke R, Kohler M, Kühn P, et al., Use of near-infrared
27 spectroscopy to assess phosphorus fractions of different plant availability in forest soils.
28 *Biogeosciences*, 2015; 12: 3415-3428. [10.5194/bg-12-3415-2015](https://doi.org/10.5194/bg-12-3415-2015).
- 29 [48] Hall JW, McNeil B, Rollins MJ, Draper I, Thompson BG, MacAloney G, Near-Infrared Spectroscopic
30 Determination of Acetate, Ammonium, Biomass, and Glycerol in an Industrial *Escherichia Coli*
31 Fermentation. *Appl. Spectrosc.*, 1996; 50: 102-108. [10.1366/0003702963906726](https://doi.org/10.1366/0003702963906726).
- 32 [49] Raju CS, Løkke MM, Sutaryo S, Ward AJ, Møller HB, NIR Monitoring of Ammonia in Anaerobic
33 Digesters Using a Diffuse Reflectance Probe. *Sensors (Basel, Switzerland)*, 2012; 12: 2340-2350.
34 [10.3390/s120202340](https://doi.org/10.3390/s120202340).
- 35 [50] Introduction to the Theory and Instrumentation for Vibrational Spectroscopy. *Handbook of*
36 *Vibrational Spectroscopy*.
- 37 [51] Rinnan A, Berg Fvd, Engelsen SB, Review of the most common pre-processing techniques for near-
38 infrared spectra. *TrAC Trends in Analytical Chemistry*, 2009; 28: 1201-1222.
39 <http://dx.doi.org/10.1016/j.trac.2009.07.007>.
- 40 [52] Carter SG, Karl DW, Inorganic phosphate assay with malachite green: An improvement and
41 evaluation. *Journal of Biochemical and Biophysical Methods*, 1982; 7: 7-13.
42 [https://doi.org/10.1016/0165-022X\(82\)90031-8](https://doi.org/10.1016/0165-022X(82)90031-8).
- 43 [53] D'Angelo E, Crutchfield J, Vandiviere M, Rapid, Sensitive, Microscale Determination of Phosphate in
44 Water and Soil Kentucky Agricultural Experiment Station Journal Series no. 00-06-165. *Journal of*
45 *Environmental Quality*, 2001; 30: 2206-2209. [10.2134/jeq2001.2206](https://doi.org/10.2134/jeq2001.2206).
- 46 [54] Feng J, Chen Y, Pu J, Yang X, Zhang C, Zhu S, et al., An improved malachite green assay of
47 phosphate: Mechanism and application. *Analytical Biochemistry*, 2011; 409: 144-149.
48 <https://doi.org/10.1016/j.ab.2010.10.025>.
- 49 [55] Itaya K, Ui M, A new micromethod for the colorimetric determination of inorganic phosphate.
50 *Clinica Chimica Acta*, 1966; 14: 361-366. [https://doi.org/10.1016/0009-8981\(66\)90114-8](https://doi.org/10.1016/0009-8981(66)90114-8).

- 1 [56] Moslen MT, Kanz MF, Ferguson AE, A stable colorimetric assay to measure toxin elevation of
2 inorganic phosphate in bile. *Analytical Biochemistry*, 1988; 168: 405-410.
3 [https://doi.org/10.1016/0003-2697\(88\)90336-3](https://doi.org/10.1016/0003-2697(88)90336-3).
- 4 [57] Penney CL, A simple micro-assay for inorganic phosphate. *Analytical Biochemistry*, 1976; 75: 201-
5 210. [https://doi.org/10.1016/0003-2697\(76\)90071-3](https://doi.org/10.1016/0003-2697(76)90071-3).
- 6 [58] Petitou M, Tuy F, Rosenfeld C, A simplified procedure for organic phosphorus determination from
7 phospholipids. *Analytical Biochemistry*, 1978; 91: 350-353. [https://doi.org/10.1016/0003-
8 2697\(78\)90849-7](https://doi.org/10.1016/0003-2697(78)90849-7).
- 9 [59] Terry Jenkins W, Marshall MM, A modified direct phosphate assay for studying ATPases. *Analytical
10 Biochemistry*, 1984; 141: 155-160. [https://doi.org/10.1016/0003-2697\(84\)90439-1](https://doi.org/10.1016/0003-2697(84)90439-1).
- 11 [60] Barberis E, Ajmone-Marsan F, Arduino E, Determination of phosphate in solution at different ionic
12 composition using malachite green. *Communications in Soil Science and Plant Analysis*, 1998; 29:
13 1167-1175. 10.1080/00103629809370017.
- 14 [61] Andersen CM, Bro R, Quantification and handling of sampling errors in instrumental measurements:
15 a case study. *Chemometrics and Intelligent Laboratory Systems*, 2004; 72: 43-50.
16 <https://doi.org/10.1016/j.chemolab.2003.12.014>.
- 17 [62] Booksh KS, Kowalski BR, Theory of Analytical Chemistry. *Analytical Chemistry*, 1994; 66: 782A-791A.
18 10.1021/ac00087a718.
- 19 [63] Lorber A, Faber K, Kowalski BR, Net Analyte Signal Calculation in Multivariate Calibration. *Analytical
20 Chemistry*, 1997; 69: 1620-1626. 10.1021/ac960862b.

21

1 **FIGURE CAPTIONS**

2 **Fig. 1** DoE representation. The samples are represented by their number in a position relative to
3 their concentration in phosphates and ammonium (expressed in mM). Samples represented in a
4 circle were selected to be part of validation set.

5

6 **Fig. 2** Raw Infrared spectra of (A) water, (D) minimal medium and (G) complex medium solutions.
7 Insets of the most interesting region for phosphate (B, E, H) and ammonium (C, F, I), on the right.
8 Assignment of the main signals for phosphates and ammonium has been included on the insets B
9 and C [38, 41].

10

11 **Fig. 3** PLS models on complex medium, whole IR region ($800 - 2250\text{ cm}^{-1}$). A) regression model
12 on the whole region for the amount of phosphates, with (B) corresponding VIP values and (C)
13 regression coefficients. D) regression model on the whole spectrum for the amount of ammonium
14 with (E) corresponding VIP values and (F) regression coefficients. The spectral region correlated to
15 the analyte, found by visual inspection and according to literature, is highlighted in blue (panels B,
16 C, E, F): these regions will be used in the following Interval models.

17

18 **Fig. 4** Comparison of all PLS models on IR (A, B and E) and NIR (C and D) data by their relative
19 Prediction Error in %. Panels A and B: models for phosphates and ammonium on IR data, both on
20 the whole spectra ($800 - 2250\text{ cm}^{-1}$) and on the intervals ($950 - 1200\text{ cm}^{-1}$ for phosphates and 1360
21 $- 1540\text{ cm}^{-1}$ for ammonium). Panels C and D: models for phosphates and ammonium on NIR data,
22 both on the whole spectra ($400 - 2500\text{ nm}$) and on the intervals ($1600-1900\text{ nm}$ for phosphates and

1 2050-2300 nm for ammonium). Panel E: models for phosphates and ammonium built on the three
2 sets of samples both on the whole spectra ($800 - 2250 \text{ cm}^{-1}$) and on the intervals ($950 - 1200 \text{ cm}^{-1}$
3 for phosphates and $1360 - 1540 \text{ cm}^{-1}$ for ammonium).

4

5 **Fig. 5** PLS models on complex medium, whole NIR region (400-3500 nm). A) regression model on
6 the whole region for the amount of phosphates, with (B) corresponding VIP values and (C)
7 regression coefficients. D) regression model on the whole spectrum for the amount of ammonium
8 with (E) corresponding VIP values and (F) regression coefficients. The spectral region correlated to
9 the analyte according to literature is highlighted panels B, C, E, F: this region will be used in the
10 following Interval model.

11

12 **Fig. 6** PLS models on three sets of samples, whole IR region ($800 - 2250 \text{ cm}^{-1}$). A) regression
13 model on the whole region for the amount of phosphates, with (B) corresponding VIP values and
14 (C) regression coefficients. D) regression model on the whole spectrum for the amount of
15 ammonium with (E) corresponding VIP values and (F) regression coefficients. The spectral region
16 correlated to the analyte, found by visual inspection and according to literature, is highlighted
17 panels B, C, E, F: this region will be used in the following Interval model.

18

19 **Fig. A1:** Comparison of the reference analysis and theoretical values. Comparison of the theoretical
20 values of phosphates with the colorimetric analysis values (a) and with the ^{31}P NMR values (b) in
21 the three, different media under study: water, minimal and complex medium. Comparison of the
22 theoretical values of ammonium with the colorimetric analysis values (c).

23

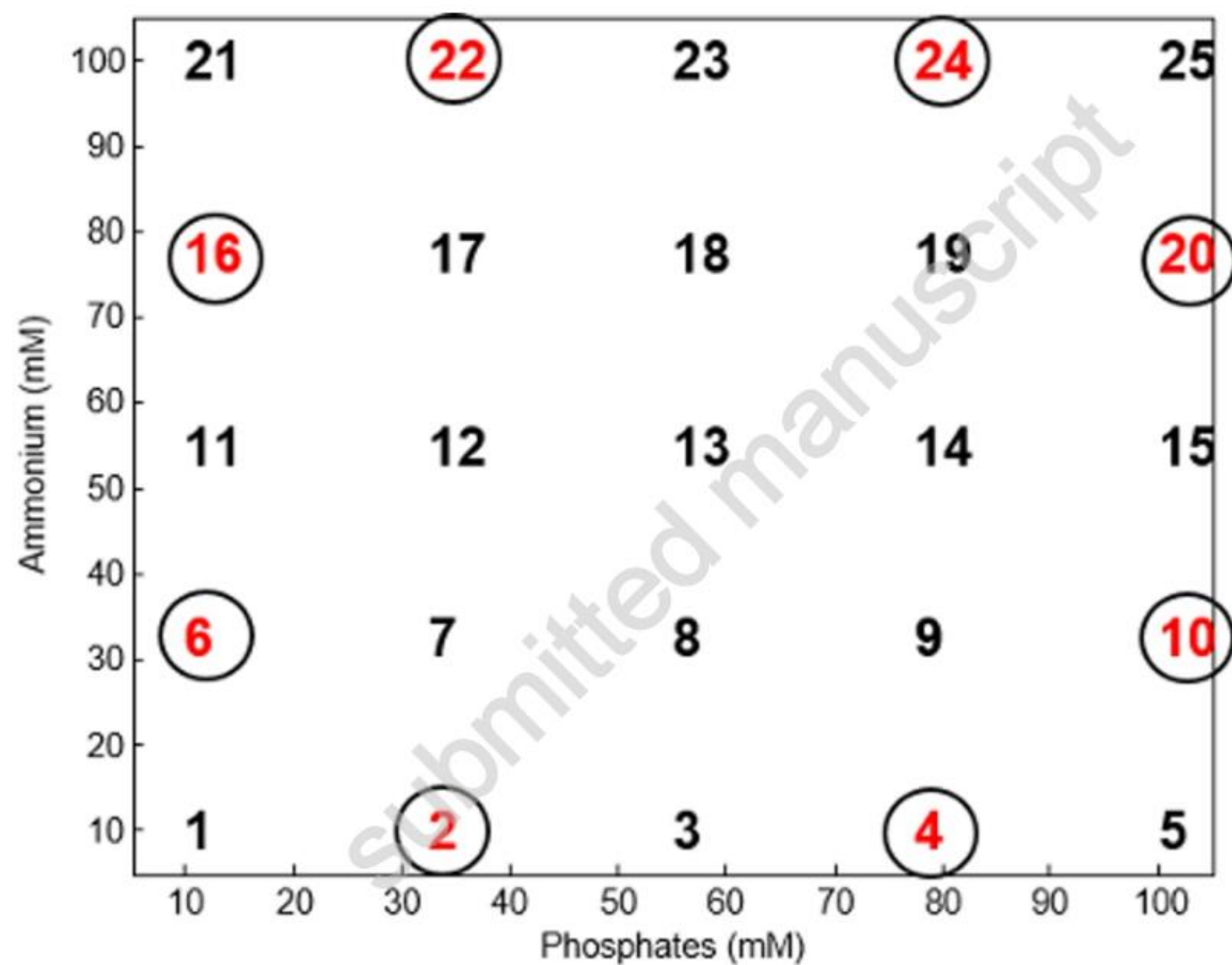
1 **Fig. B1:** Contribution to the Prediction Error given by the Error in X (spectra) and y (reference
2 measurements), for both IR and NIR spectra and for both analytes.

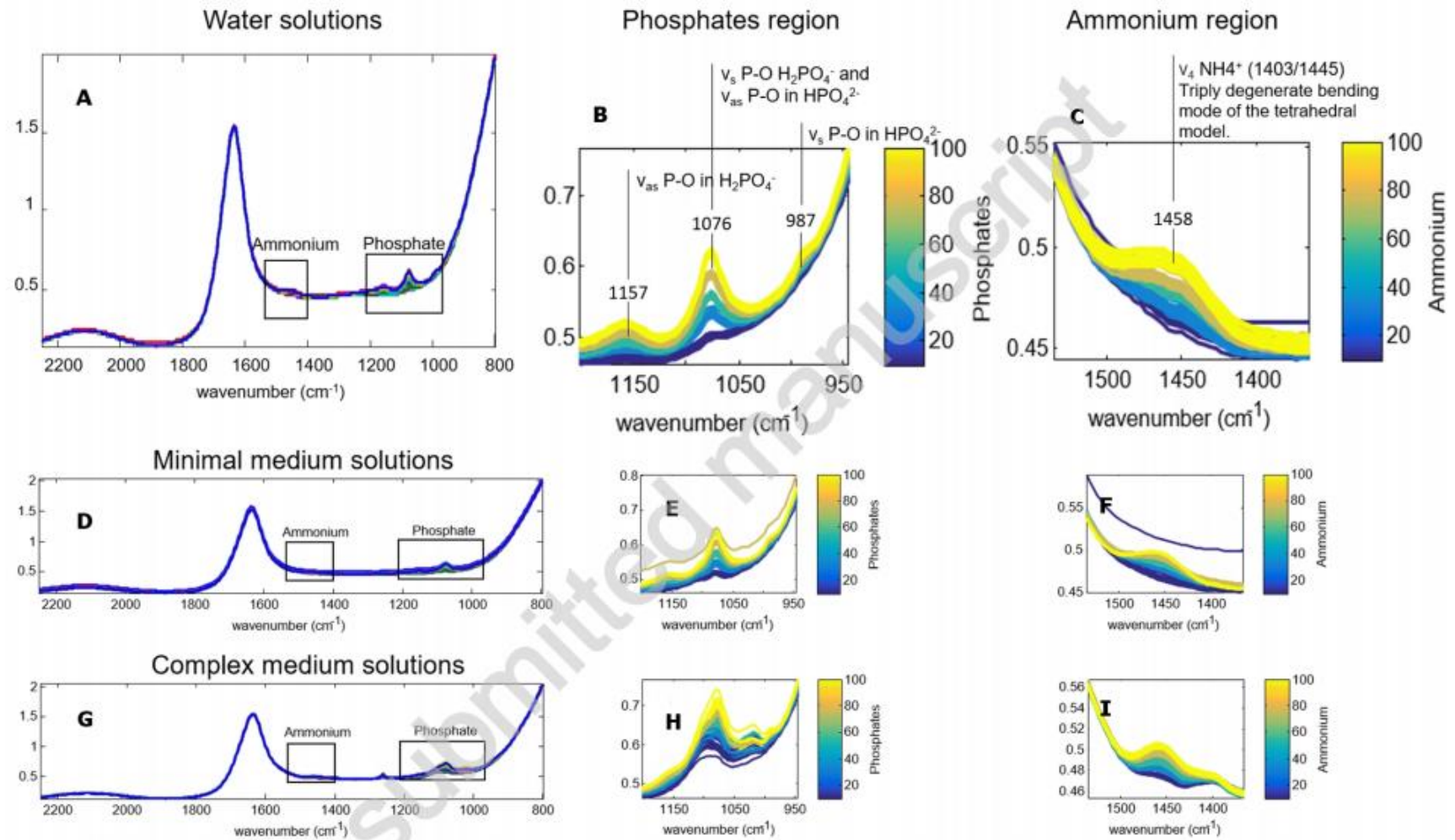
3

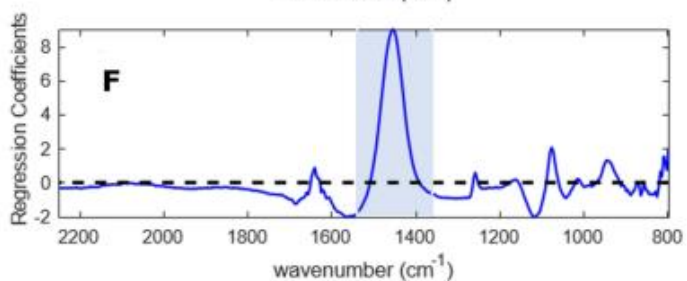
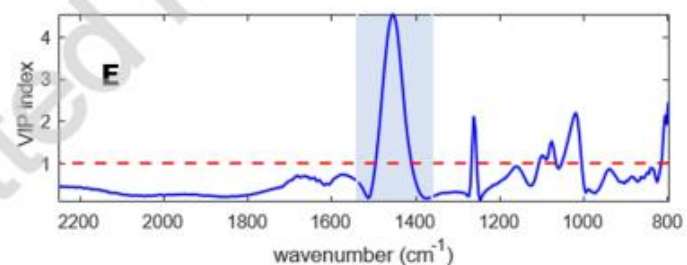
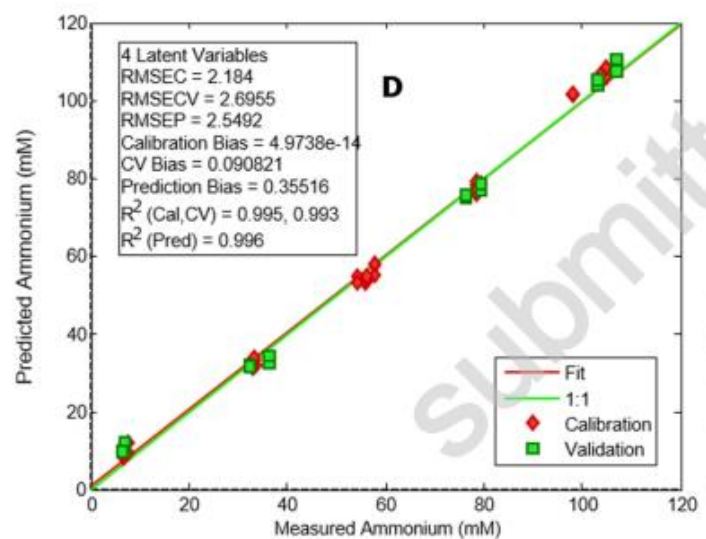
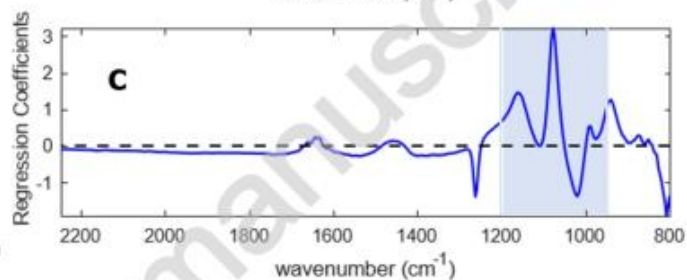
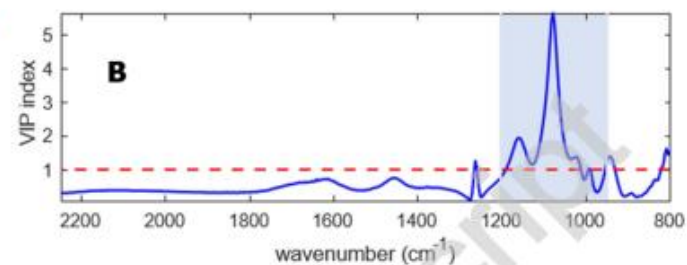
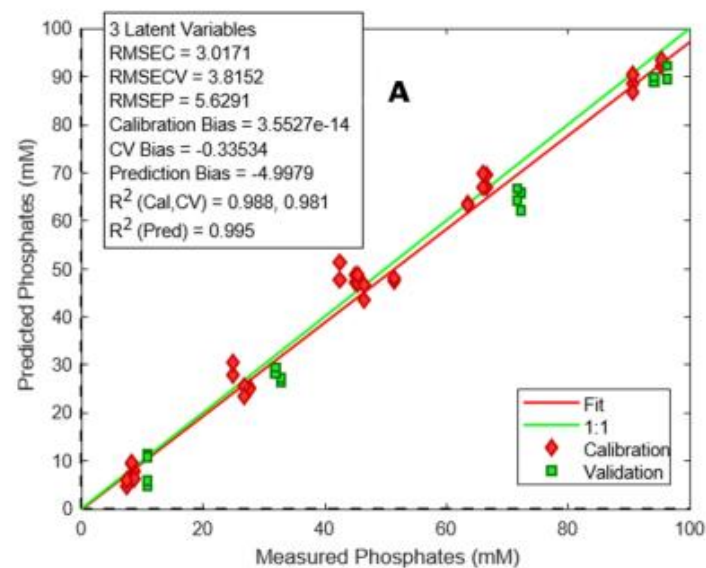
4 **Fig. B2:** Interpretation of the Error in X (spectra) by its contributions: Variance in X and norm of b .

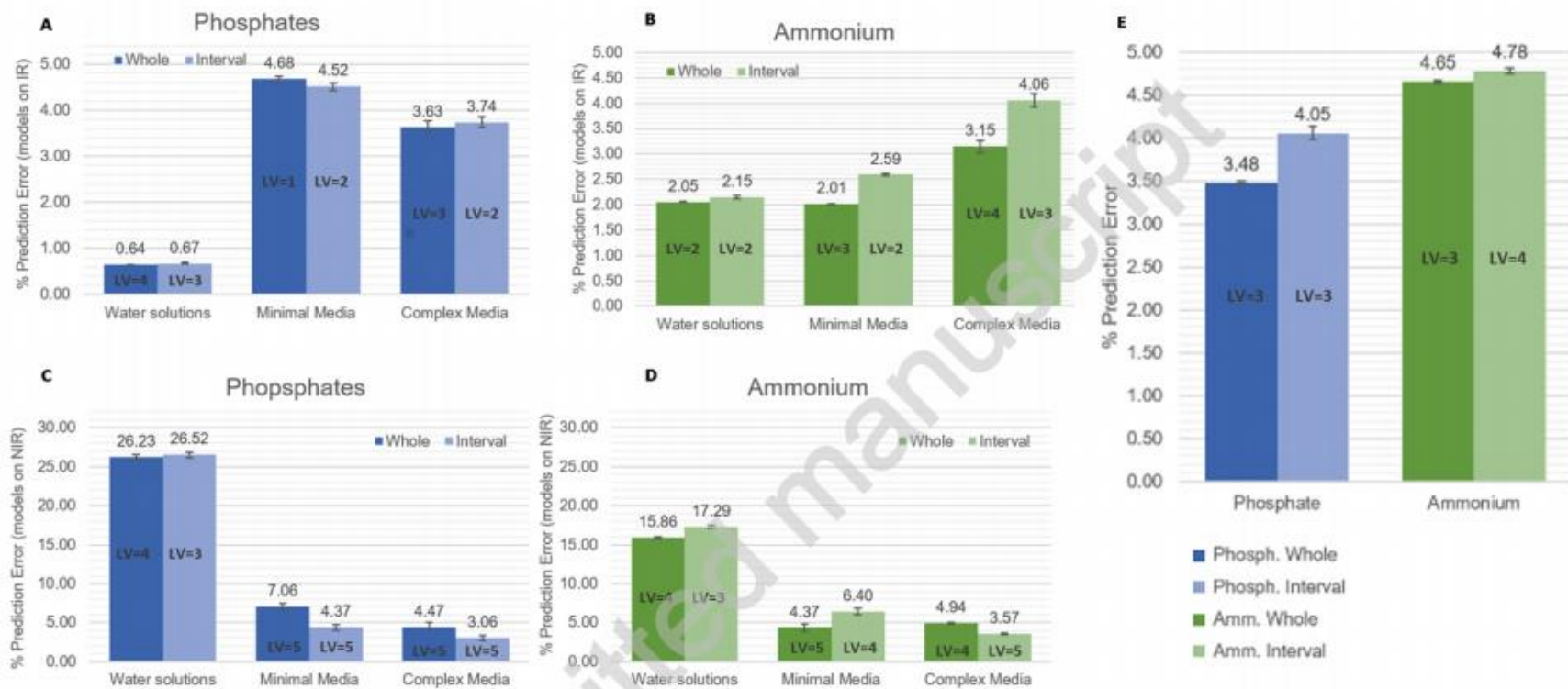
5

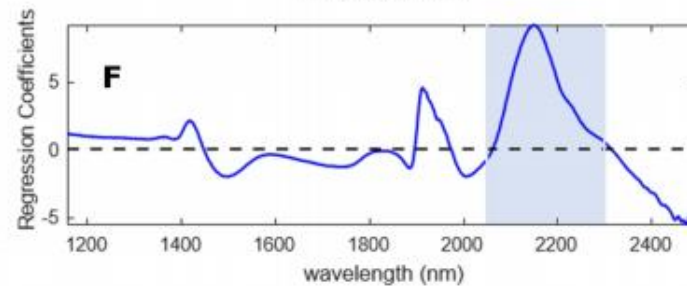
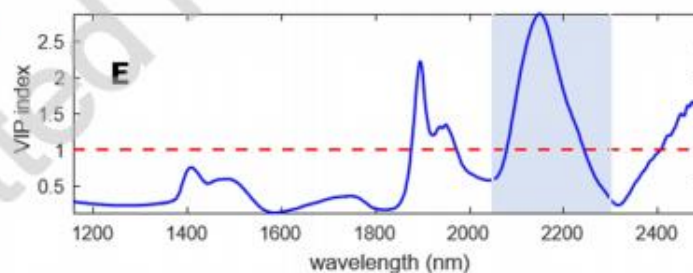
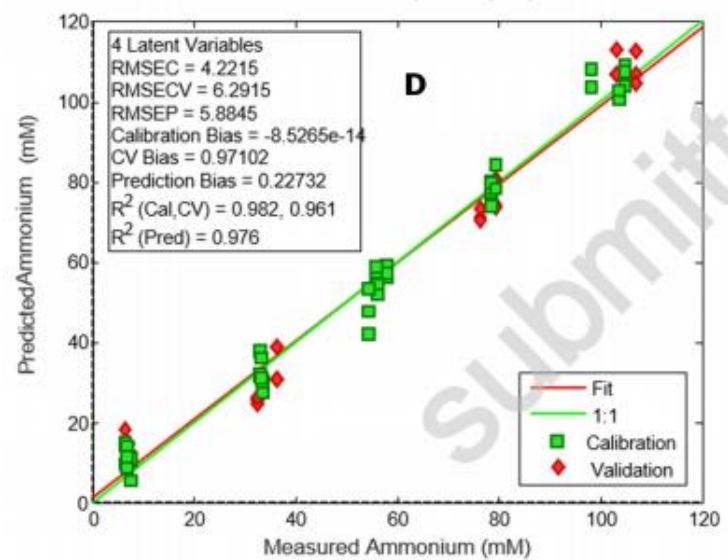
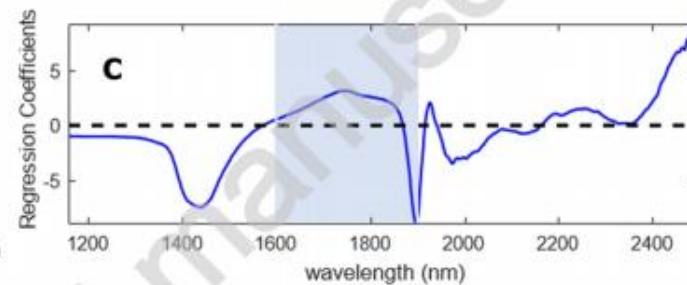
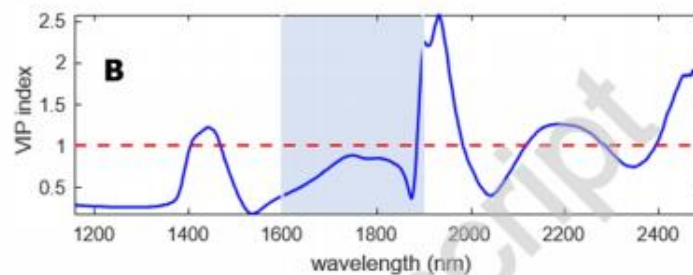
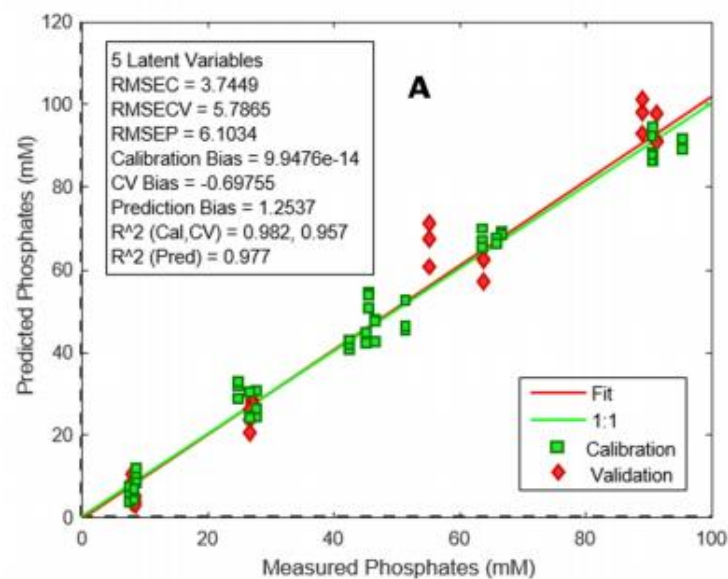
Submitted manuscript

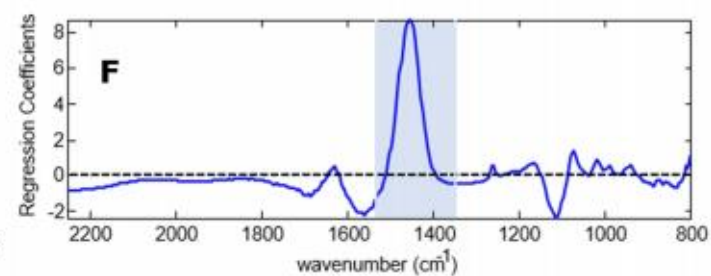
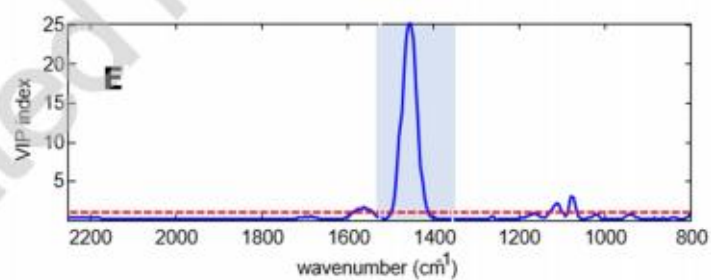
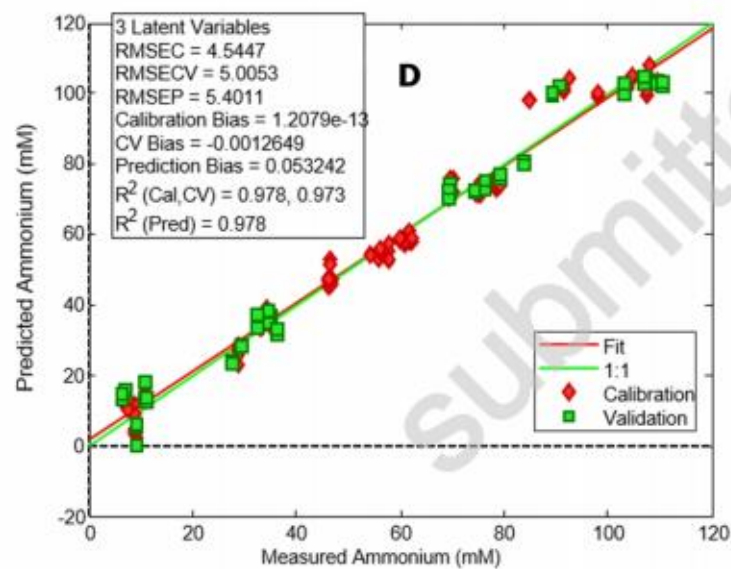
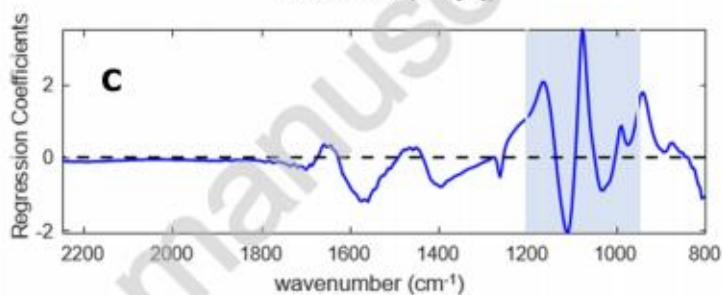
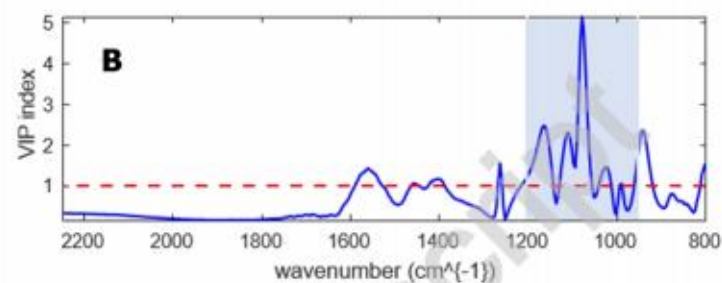
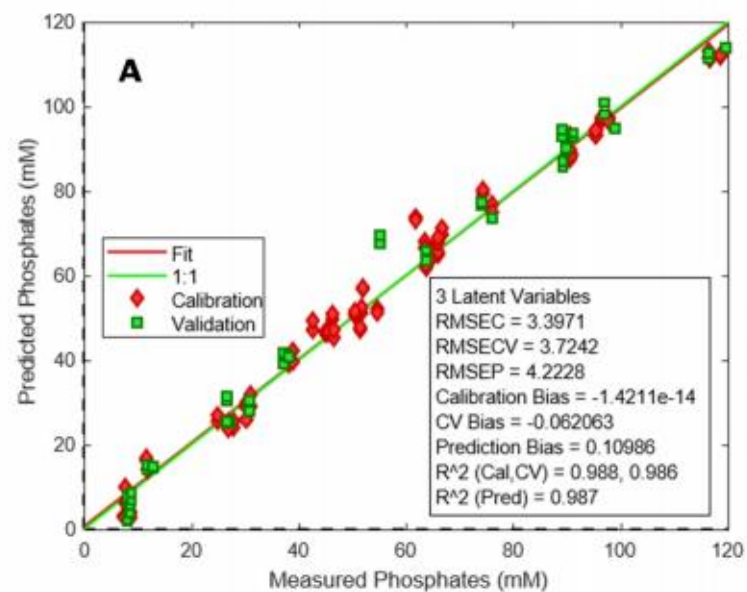


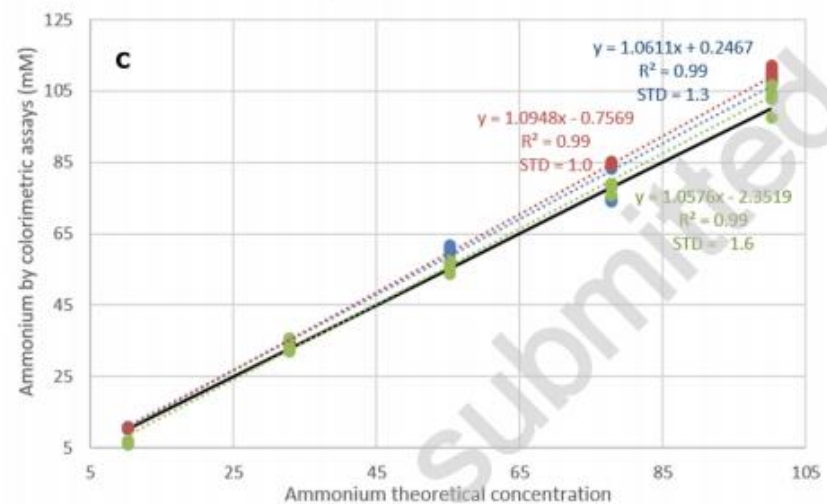
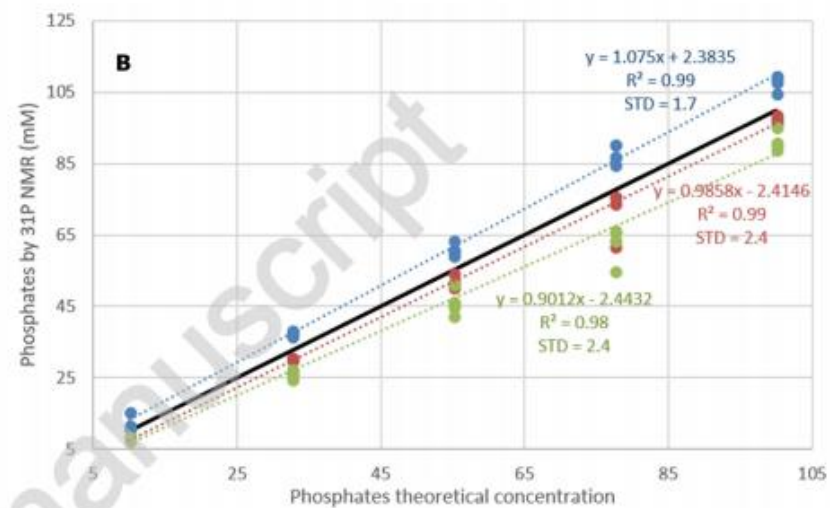
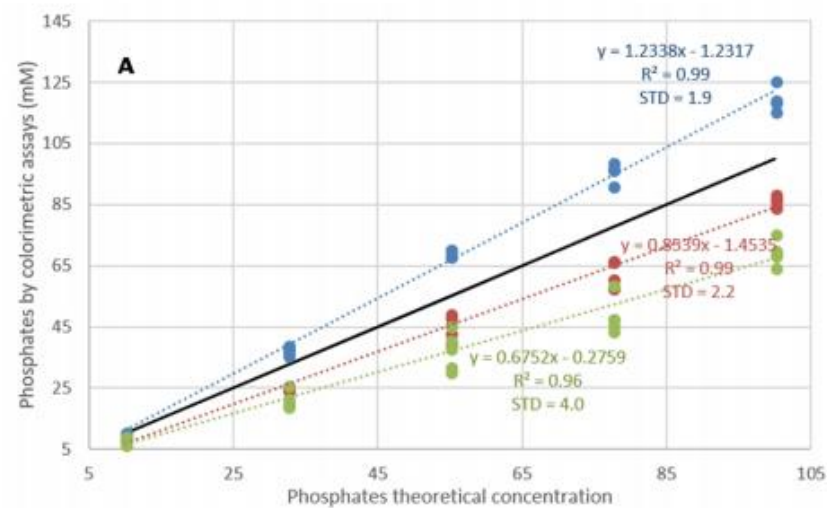




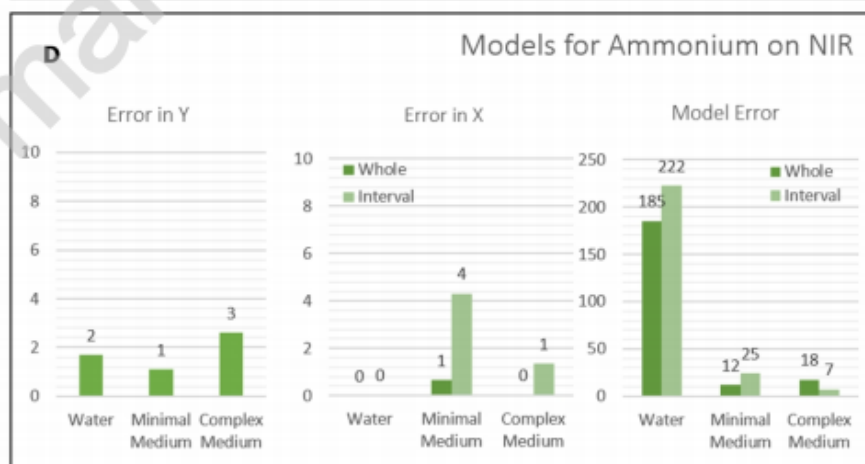
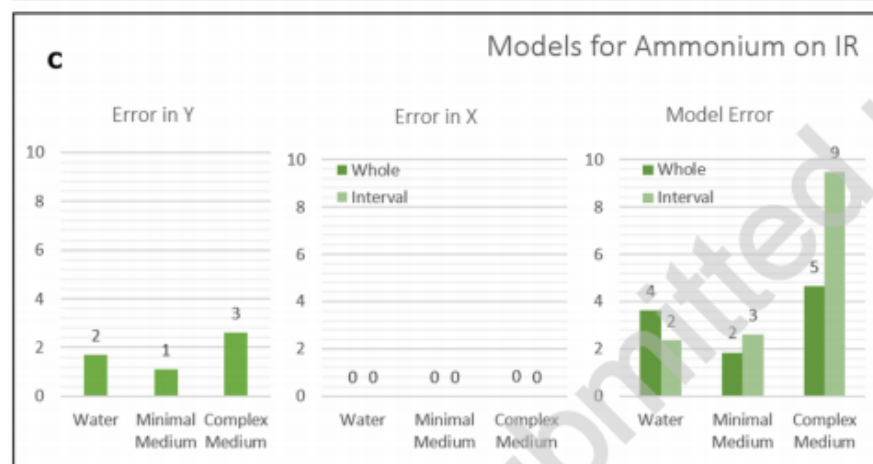
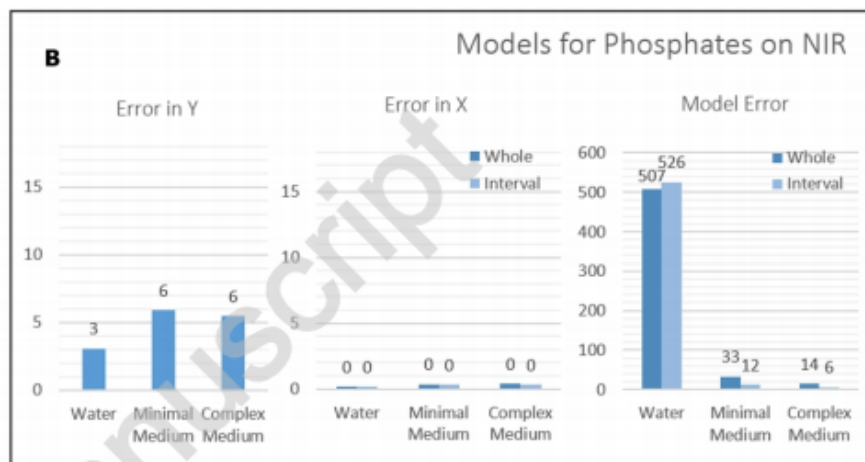
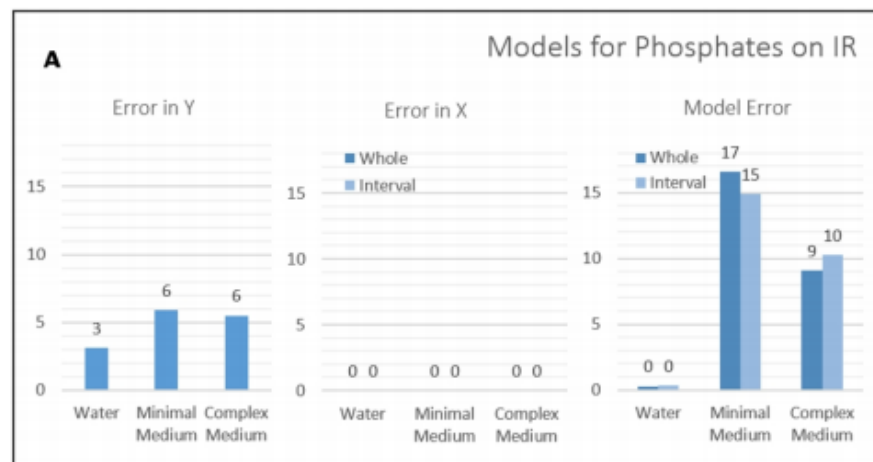


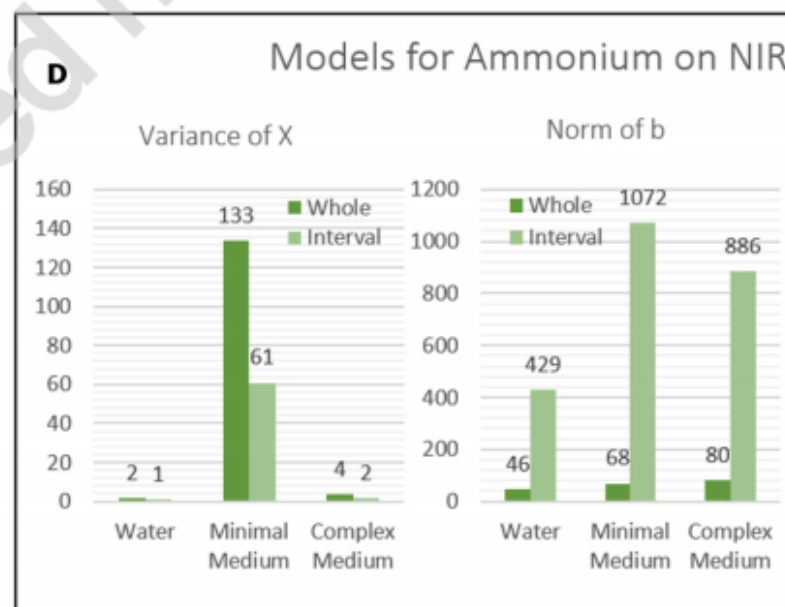
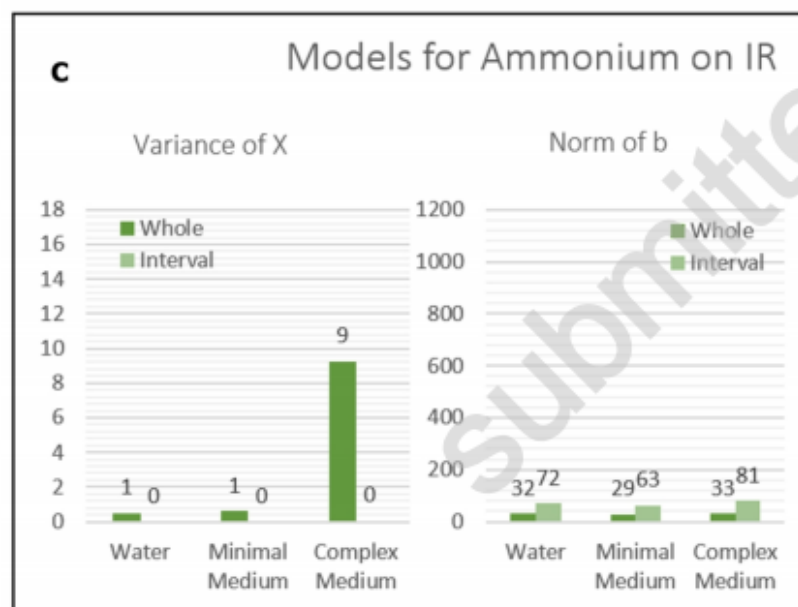
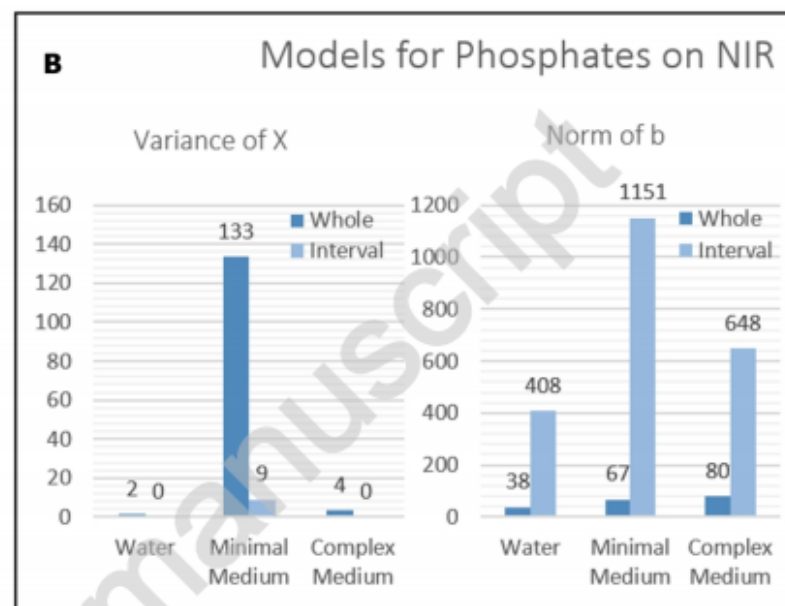
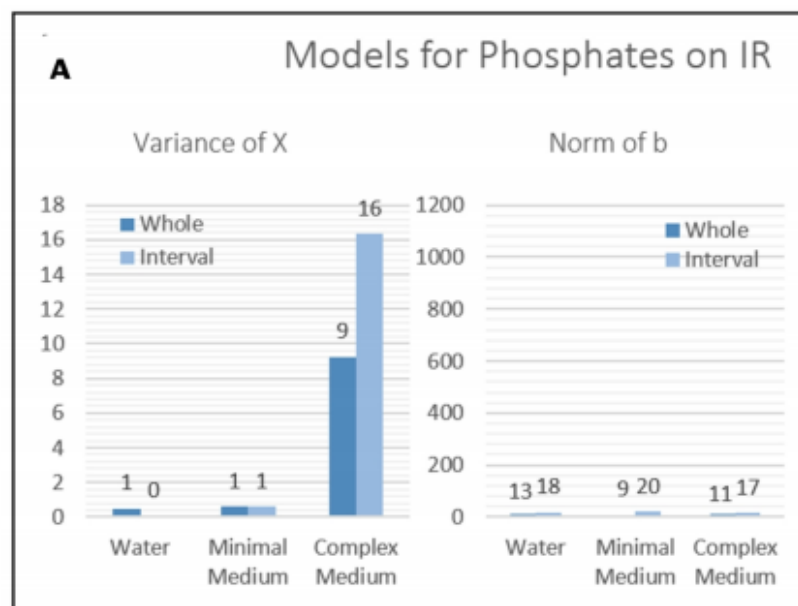






- Water
- Minimal medium
- Complex medium
- Theoretical
- Linear (Water)
- Linear (Minimal medium)
- Linear (Complex medium)





AUTHOR DECLARATION

We wish to confirm that there are no known conflicts of interest associated with this publication and that the funding source supporting this work has not influenced its outcome in any way.

We confirm that the manuscript has been read and approved by all named authors and that there are no other persons who satisfied the criteria for authorship but are not listed. We further confirm that the order of authors listed in the manuscript has been approved by all of us.

We declare that all the authors have materially participated in the research and/or article preparation, and in detail, we summarize the individual authors contributions below:

- Katrin Pontius: media preparation and reference analysis with the assays, writing parts of the manuscript, active revision of the manuscript, suggesting reviewers and journal.
- Giulia Praticò: sample preparation and reference analysis with the NMR, writing parts of the manuscript, active revision of the manuscript.
- Flemming H. Larsen: help with the reference analysis with the NMR, writing parts of the manuscript, active revision of the manuscript.
- Thomas Skov: help in structuring the manuscript and feedback on it.
- Nils Arneborg: help in structuring the manuscript and feedback on it.
- Anna Eliasson Lantz: active participation in project related meetings, feedback on the manuscript.
- Marta Bevilacqua: sample preparation and analysis with vibrational spectroscopies, planning the experimental design and carrying out of the chemometric analysis, writing main parts of the manuscript, revision of the manuscript, submission and related duties.

We confirm that we have given due consideration to the protection of intellectual property associated with this work and that there are no impediments to publication, including the timing of publication, with respect to intellectual property. In so doing we confirm that we have followed the regulations of our institutions concerning intellectual property.

We understand that the Corresponding Author is the sole contact for the Editorial process (including Editorial Manager and direct communications with the office). She is responsible for communicating with the other authors about progress, submissions of revisions and final approval of proofs. We confirm that we have provided a current, correct email address which is accessible by the Corresponding Author and which has been configured to accept email from (marta@food.ku.dk)

Signed by all authors as follows:

AUTHOR:	DATE:	SIGNATURE:
Katrin Pontius.....	4/2, 2019	Katrin Pontius
Giulia Praticò.....	4/2/2019	Giulia Praticò
Flemming H. Larsen.....	4/2 - 2019	Flemming H. Larsen
Thomas Skov.....	3/2 - 2019	Thomas Skov
Nils Arneborg.....	4/2 - 2019	Nils Arneborg
Anna Eliasson Lantz.....	4/2 - 2019	Anna Eliasson Lantz
Marta Bevilacqua.....	4/2 - 2019	Marta Bevilacqua

Supplementary Material

SUPPLEMENTARY TABLES

Table S1. Concentrations of the synthetic samples prepared. Comparison between theoretical and reference values (measured by colorimetric assay kits) for each sample. Samples included in the validation set are listed in red and indicated with a star on the column of their name. The standard deviation for each set of samples is calculated on the 5 replicates at each level of concentration and then averaged on the different levels.

Samples	Theoretical	Water	Minimal	Complex	Theoretical	Water	Minimal	Complex
	values	(std=1.7)	medium	medium	values	(std=1.3)	medium	medium
			(std=2.4)	(std=2.4)			(std=1.0)	(std=1.6)
1	10	15.8	7.7	8.7	10	10.9	11.1	7.6
2*	32.5	38.3	30.8	27.2	10	10.8	11.4	7.0
3	55	61.0	54.7	51.4	10	10.9	11.3	6.4
4*	77.5	90.5	76.1	63.8	10	11.1	11.3	6.5
5	100	109.8	97.7	95.4	10	10.9	11.0	6.7
6*	10	11.5	8.1	8.0	32.5	35.2	35.8	36.3
7	32.5	37.9	30.2	24.9	32.5	34.5	35.0	33.6
8	55	60.7	51.7	45.2	32.5	34.6	35.3	33.1
9	77.5	87.0	76.1	66.7	32.5	35.1	35.1	33.4
10*	100	104.8	98.9	91.3	32.5	34.7	34.0	32.4
11	10	12.0	8.1	7.5	55	61.8	56.3	58.0
12	32.5	38.5	30.2	27.8	55	60.6	56.8	58.0
13	55	63.8	46.2	42.6	55	62.3	56.9	55.9

14	77.5	90.7	61.9	66.0	55	61.1	57.0	56.3
15	100	109.7	96.6	90.7	55	60.1	56.2	54.4
16*	10	11.1	8.8	8.6	77.5	74.6	85.0	79.4
17	32.5	36.8	30.9	26.8	77.5	75.7	85.7	78.7
18	55	60.8	50.5	45.6	77.5	75.3	84.6	78.7
19	77.5	87.3	74.4	63.6	77.5	74.9	85.0	79.4
20*	100	108.6	97.1	89.1	77.5	83.9	84.6	76.5
21	10	11.1	7.8	8.3	100	108.2	111.5	104.8
22*	32.5	37.9	26.6	26.6	100	109.7	110.5	103.3
23	55	59.5	52.1	46.6	100	109.7	112.7	103.7
24*	77.5	84.8	74.1	55.2	100	110.8	108.8	107.2
25	100	108.0	97.8	90.6	100	107.8	103.5	98.2

Table S2. Contribution to the Prediction Error given by the Error in X (spectra) in y (reference measurements) and in the model itself.

			Water		Minimal Medium		Complex Medium	
			Phosp.	Amm.	Phosp.	Amm.	Phosp.	Amm.
Error in y [mM ²]			3.1	1.7	5.9	1.1	5.5	2.6
IR	Error in X	Whole	0.0	0.0	0.0	0.0	0.0	0.0
	[mM ²]	Interval	0.0	0.0	0.0	0.0	0.0	0.0
	Model Error	Whole	0.3	3.6	16.6	1.9	9.1	4.6
	[mM ²]	Interval	0.3	2.4	14.9	2.6	10.2	9.5
NIR	Error in X	Whole	0.2	0.0	0.4	0.7	0.4	0.0
	[mM ²]	Interval	0.2	0.0	0.4	4.3	0.4	1.4
	Model Error	Whole	507.1	185.2	33.3	12	14.0	17.69
	[mM ²]	Interval	525.8	222.2	12.4	24.60	6.1	7.50

Table S3. Interpretation of the Error in X (spectra) by its contributions: Variance in X and norm of b.

			Water		Minimal Medium		Complex Medium	
			Phosp.	Amm.	Phosp.	Amm.	Phosp.	Amm.
IR	Variance of X	Whole	0.5	0.5	0.6	0.6	9.3	9.3
	($\times 10^{-6}$)	Interval	0.1	0.0	0.6	0.0	16.4	0.1
	Norm of b	Whole	12.7	31.7	8.9	29.4	11.1	32.8
		Interval	18.5	72.2	19.6	62.9	16.6	81.2
NIR	Variance of X	Whole	1.5	1.5	133.5	133.5	3.6	3.6
	($\times 10^{-6}$)	Interval	0.4	1.1	9.0	60.7	0.3	1.7
	Norm of b	Whole	38.1	46.0	67.2	68.0	78.9	79.9
		Interval	407.8	429.1	1151.5	1071.9	648.3	885.8

Table S4. PLS models on IR data. Comparison of all PLS models on phosphates and ammonium, both on the whole spectra (800 – 2250 cm^{-1}) and on the intervals (950 – 1200 cm^{-1} for phosphates; 1360 – 1540 cm^{-1} for ammonium).

Model			#	R^2	Calibration		CV		Prediction	
			LV (pred.)		RMSE	Bias	RMSE	Bias	RMSE	Bias
					(mM)	(mM)	(mM)	(mM)	(mM)	(mM)
Water	Phosphates	Whole	4	1	0.54	0	0.85	-0.07	1.44	0.3
		Interval	3	1	0.57	0	0.78	0.02	0.91	0.3
	Ammonium	Whole	2	1	1.90	0	2.34	0.2	1.66	0.2
		Interval	2	1	1.54	0	2.03	0.1	1.74	0.7
Minimal media	Phosphates	Whole	1	0.99	4.07	0	4.28	-0.08	3.12	-1.1
		Interval	2	1	3.86	0	4.05	-0.2	1.65	-0.99
	Ammonium	Whole	3	1	1.36	0	1.64	0.01	1.63	-0.6
		Interval	2	1	1.62	0	1.71	0.03	2.10	-0.3
Complex media	Phosphates	Whole	3	0.99	3.02	0	3.81	-0.3	5.62	-5.0
		Interval	2	1	3.20	0	4.23	-0.6	5.93	-5.4
	Ammonium	Whole	4	1	2.18	0	2.70	0.09	2.55	0.4
		Interval	3	0.99	3.09	0	3.38	0.06	3.29	0.9

Table S5. PLS models on NIR data. Comparison of all PLS models on phosphates and ammonium, both on the whole spectra (400-2500 nm) and on the intervals (1600-1900 nm for phosphates; 2050-2300 nm for ammonium).

Model			#	R ²	Calibration	CV		Prediction			
					LV (pred.)	RMSE	Bias	RMSE	Bias	RMSE	Bias
						(mM)	(mM)	(mM)	(mM)	(mM)	(mM)
Water	Phosphates	Whole	4	0.40	22.52	0.0	26.12	-0.1	32.0	-3.0	
		Interval	3	0.51	22.93	0.0	26.19	-0.6	29.4	-3.7	
	Ammonium	Whole	4	0.96	13.61	0.0	14.86	0.5	8.66	2.0	
		Interval	3	0.97	14.91	0.0	17.08	0.5	7.92	1.6	
Minimal media	Phosphates	Whole	5	0.98	5.77	0.0	7.01	0.4	4.9	0.3	
		Interval	5	0.99	3.58	0.0	4.30	0.2	2.8	-1.0	
	Ammonium	Whole	5	0.99	3.57	0.0	5.53	0.8	4.04	-0.4	
		Interval	4	0.97	5.38	0.0	6.28	0.1	7.04	-0.3	
Complex media	Phosphates	Whole	5	0.98	3.74	0.0	5.79	-0.7	6.1	1.2	
		Interval	5	0.99	2.54	0.0	3.55	-0.4	3.9	0.5	
	Ammonium	Whole	4	0.98	4.22	0.0	6.29	1.0	5.88	0.2	
		Interval	5	0.99	2.99	0.0	5.05	0.8	3.07	0.8	

Table S6. Comparison of the best IR and NIR models chosen among Whole (w) and Interval (i) models, based on the lowest prediction errors)

Model			#	R ²	Calibration		CV		Prediction	
					LV (pred.)					
					RMSE	Bias	RMSE	Bias	RMSE	Bias
					(mM)	(mM)	(mM)	(mM)	(mM)	(mM)
Water	Phosphates	IR (w)	4	1	0.54	0	0.85	-0.07	1.44	0.3
		NIR (w)	4	0.40	22.52	0.0	26.12	-0.1	32.0	-3.0
	Ammonium	IR (w)	2	1	1.90	0	2.34	0.2	1.66	0.2
		NIR (w)	4	0.96	13.61	0.0	14.86	0.5	8.66	2.0
Minimal media	Phosphates	IR (i)	2	1	3.86	0	4.05	-0.2	1.65	-0.99
		NIR (i)	5	0.99	3.58	0.0	4.30	0.2	2.8	-1.0
	Ammonium	IR (w)	3	1	1.36	0	1.64	0.01	1.63	-0.6
		NIR (w)	5	0.99	3.57	0.0	5.53	0.8	4.04	-0.4
Complex media	Phosphates	IR (w)	3	0.99	3.02	0	3.81	-0.3	5.62	-5.0
		NIR (i)	5	0.99	2.54	0.0	3.55	-0.4	3.9	0.5
	Ammonium	IR (w)	4	1	2.18	0	2.70	0.09	2.55	0.4
		NIR (i)	5	0.99	2.99	0.0	5.05	0.8	3.07	0.8

Table S7. Models built on the three sets of samples all together. Comparison of PLS models on phosphates and ammonium, both on the whole spectra (800 – 2250 cm^{-1}) and on the intervals (950 – 1200 cm^{-1} for phosphates; 1360 – 1540 cm^{-1} for ammonium).

Model		#	R^2	Calibration		CV		Prediction	
				RMSE	Bias	RMSE	Bias	RMSE	Bias
		LV	(pred.)	(mM)	(mM)	(mM)	(mM)	(mM)	(mM)
Phosphates	Whole	3	0.99	3.40	0	3.72	-0.06	4.22	0.1
	Interval	3	0.99	3.96	0	4.15	-0.05	4.49	-0.05
Ammonium	Whole	3	0.98	4.54	0	5.00	-0.001	5.40	0.05
	Interval	4	0.97	4.66	0	5.25	0.03	6.04	0.6

SUPPLEMENTARY FIGURES

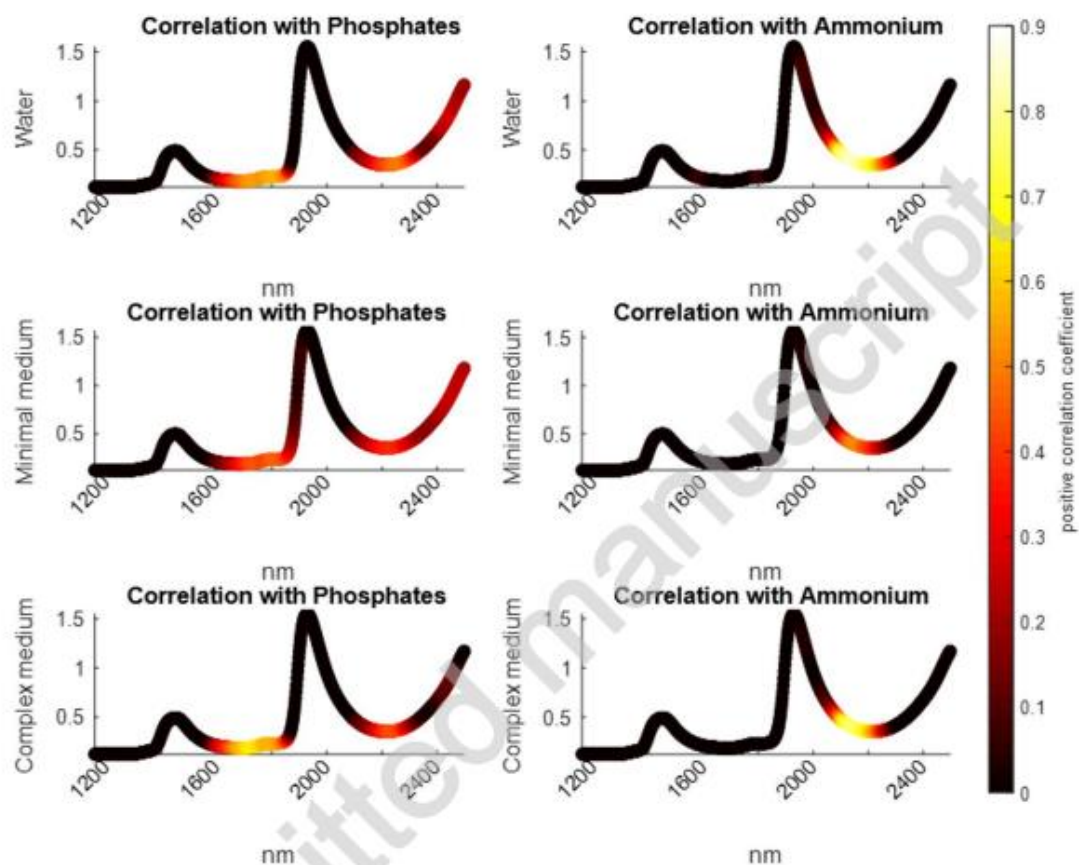


Fig. S1 Positive correlation coefficients of each wavelength in the NIR spectra with the values of phosphates and ammonium

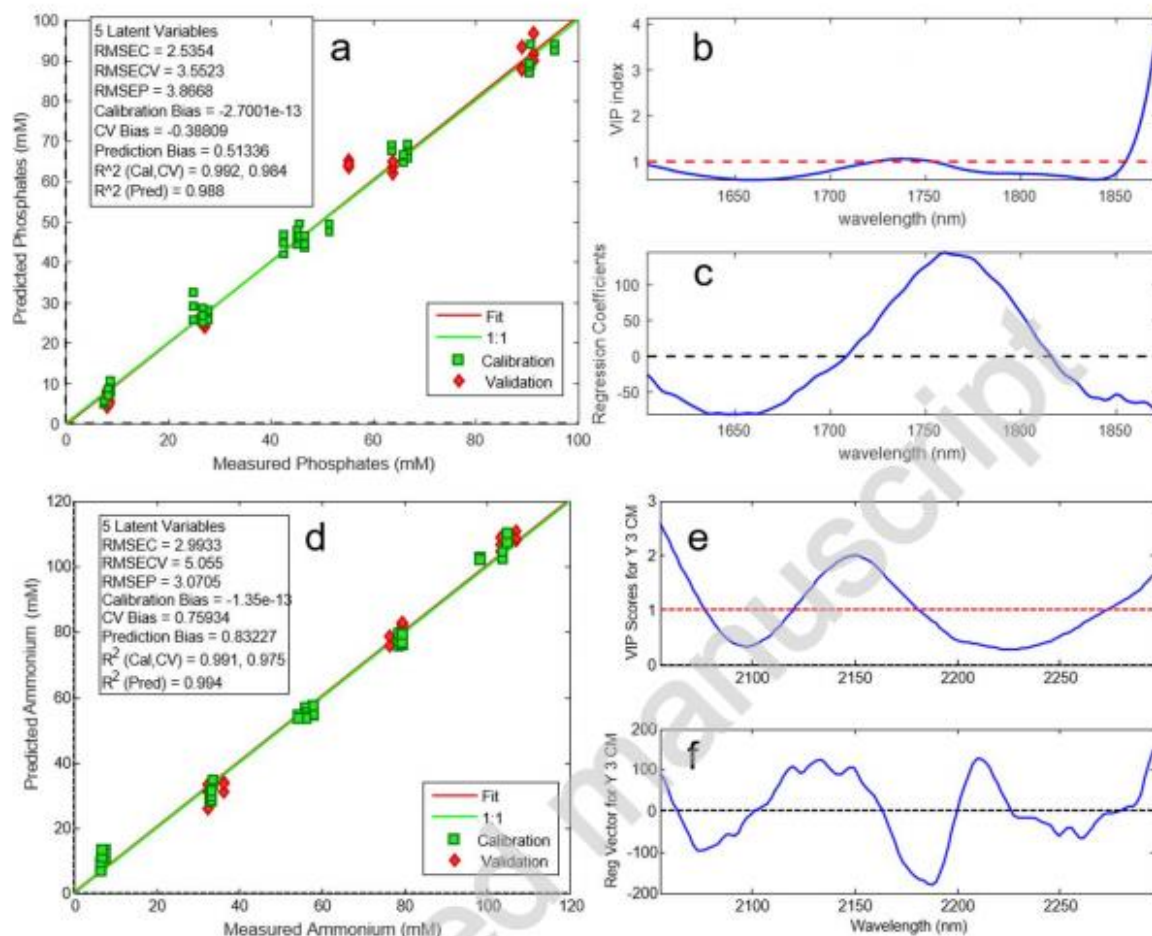


Fig. S2 PLS models on complex media, selected NIR regions. a) regression model on the whole region for the amount of phosphates, with (b) corresponding VIP values and (c) regression coefficients. d) regression model on the whole spectrum for the amount of ammonium with (e) corresponding VIP values and (f) regression coefficients.